INDUSTRIAL PSYCHOLOGY

Its Theoretical and Social Foundations

A REVISION OF
INDUSTRIAL PSYCHOLOGY
AND ITS SOCIAL FOUNDATIONS
REVISED EDITION

Milton L. Blum & James C. Naylor

NOT FOR SALE IN THE U.S. OR CANADA

This Harper International Edition is an unabridged, photo-offset reproduction of the American edition. It is for sale in Japan by JOHN WEATHERHILL, INC., 7-6-13 Roppongi, Minato-ku, Tokyo; in the United Kingdom and on the continent of Europe by HARPER & ROW, LTD., 69 Great Russell Street, London W.C. 1; and in other designated countries by HARPER & ROW, PUBLISHERS, INCORPORATED, 49 East 33rd Street, New York, N.Y. 10016 / Library of Congress catalog card number 69-16210 / Printed in Japan / First printing, June, 1968.

To Naomi and Georgia

CONTENTS

PREFACE

Industrial psychology continues to be growing and in a state of transition, and this revision is an expression of the kinds of change that are taking place. The motives for this third edition are similar in many ways to those of the earlier books. They represent the authors' views of industrial psychology at present and their thoughts about where it is going in the future. In 1956, when this book was previously revised, industrial psychology was in the process of truly entering what might be called the "Human Relations" phase of studying man and his place in the world of work. Although the foundation for this movement had occurred years earlier in the classical Western Electric research, it was not until the 1950s that industrial psychology became bold enough to join hands with social psychology and to begin a systematic look at the many interpersonal aspects of human interrelationships existing in the work situation. Indeed, some industrial psychologists did not welcome this move. For years, industrial psychology had remained a secure, scientifically respectable discipline under the guise of a strong empirical emphasis best represented by the traditional areas of employee selection and placement, performance appraisal, and job evaluation. The social aspects of the working man were generally avoided, as they did not easily lend themselves to empirical investigation and so did not fit into the scheme of things.

However, the change in emphasis did occur, and we like to think that the earlier editions of this book contributed to this change by stressing the social foundations of industrial psychology and their growing importance.

Industrial psychology in the mid 1960s seems to be experiencing a transition of emphasis just as substantial as that which occurred in the early 1950s. However, its nature is different. The change which occurred in the early 1950s was a change in the *kinds* of problems and variables considered important to advance our knowledge. The present change of emphasis is one dealing with the *way* in which industrial psychology approaches the study of man at work. It is a change in philosophy rather than a change in content.

What is this philosophical change? Quite simply, it is a growing tendency for modern industrial psychologists to become increasingly involved in the development of theory. The emphasis is changing from a philosophy of rigid empiricism applied to a variety of seemingly unrelated situations and phenomena to a philosophy concerned with the establishment of underlying behavioral principles which will permit a fuller understanding of man's behavior in realistic task environments. From a scientific viewpoint, it is a more mature and exceedingly richer philosophy than its predecessor. As authors, we believe that if industrial psychology is to evolve as a science, its theoretical basis must be developed and emphasized.

Actually, such a transition had to occur—it is inevitable in any maturing scientific area. However, it has been long overdue in industrial psychology, and this continuing lack of attention to theory has been a major source of difficulty in advancing the scope of knowledge within the field. The difficulty was most aptly stated by Dr. Brent Baxter in the closing paragraph of his presidential address to the Division of Industrial Psychology in September, 1965:

> In summary, my message urges you, each one of you, to turn some part of your attention to the development of your own broad theory of human behavior. At least

> adopt or adapt something that you put in writing and then apply it as a conscious guide in your daily activities. To find time, take some now given to the development of techniques. Spend more time determining where we are going and less on how we are going to get there. I'm not suggesting you should abandon your empiricism but rather give it some roots. As we all more consciously think of these fundamentals, we will create a body of knowledge that makes sense rather than confusion. We will create an image of a scientist rather than a technician. We will have status of which we can be proud. We will draw together rather than apart. This advice may well pertain not only to the variation among psychologists in industry but also to the many specialties within psychology. Thinking of fundamentals will serve to bring all of us closer together.

Kurt Lewin once said "There is nothing so practical as a good theory!" We firmly believe in this position and ask our readers to pause and think about how right he was. It is in this spirit that the current book was revised. The theme throughout is one of emphasizing emerging theory development in industrial psychology wherever it seems to be taking place. Some of the efforts seem to be disjointed, others incomplete, and many in conflict. Even though a unified theoretical system has not appeared, we hope that our efforts will contribute to the theoretical promise of industrial psychology.

The purpose, then, of the third edition is to mirror the present transition within industrial psychology in the same fashion as the 1956 edition mirrored the contemporary change within the field at that time. Incidentally, we did not overlook the studies and research efforts since the last edition. We hope we have selected wisely from them in those studies which we have discussed here.

A second objective of the present revision is to make the general level of the presentation somewhat more sophisticated than has been the case with earlier editions. Experience with students over the past few years has led to the conclusion that today's college student is a more knowledgeable and better trained person than the student of the past. In keeping with this increase in knowledge, it becomes, we think, necessary to upgrade the level of presentation traditionally found in the undergraduate textbook. We do not share the view that college is for illiterates and that texts and courses should be watered down.

The organization of material in this revision differs somewhat from the second edition. The classical areas of industrial psychology are presented in the early chapters: Prediction techniques (Chapter 2 and 3), industrial tests (Chapter 4), and other selection techniques (Chapter 5) are each examined in turn. In these chapters the attempt is made to present the conceptual and theoretical models and problems critical to a basic understanding of the topic. Chapter 6 is a detailed discussion of the criterion problem. This problem was deliberately given more than the usual emphasis since we feel it is one of the most important topics in industrial psychology and perhaps even in all psychology. Chapters 7, 8, and 9 discuss performance appraisal (a logical continuator of the criterion chapter), training, and attitude measurement. With this last chapter the transition begins to take place into the more social aspects of industrial psychology such as the Hawthorne studies (Chapter 10), motivation (Chapter 11), job satisfaction (Chapter 12), morale (Chapter 13), and leadership and supervision (Chapter 14). Throughout all these chapters the emphasis is one of examining the models or theories which seem to be developing in each of these areas.

Chapters 15 and 16 examine decision making and organizational behavior, two very important and dynamic areas in modern industrial psychology. Chapters 17

through 20 cover work variables such as job analysis, accidents and fatigue, the work environment, principles of human performance, and general industrial problems.

While the chapter order in a text such as this is always an arbitrary matter, the present organization of the material seems to the authors to form a logical sequence. However, the book has been written so that any chapter can be read separately or out of the present sequence.

ACKNOWLEDGMENTS

Students and clients highly deserve recognition as contributors to the thoughts expressed in this book. The challenging questions on the parts of students have resulted in the development of both hypothesis and theory.

The many authors and publishers quoted have been most kind in granting their generous permissions.

A strong vote of appreciation should go to those colleagues who were willing to read and comment critically on various portions of the text. In particular, recognition should go to Dr. Norman L. Vincent, Dr. Ralph Stogdill, Dr. Robert J. Wherry, Jr., Dr. Irwin L. Goldstein and Dr. Arthur L. Dudycha. Also to be commended are those students who served as "test subjects" for the manuscript as it underwent revision. Their feedback and comments were invaluable.

Finally, a very special debt of gratitude is owed to Mrs. Geraldine Hickman for her outstanding secretarial assistance over the entire period of the preparation of this manuscript.

MILTON L. BLUM
JAMES C. NAYLOR

INDUSTRIAL PSYCHOLOGY

INTRODUCTION 1

Psychology is the study of man with the aid of scientific methodology. It uses the experimental method—observation under controlled conditions—to gather data. It also uses other methods for data collection, such as the case history method and the longitudinal study of development as it takes place. Psychology accepts introspection (a subject's report that describes mental processes) as a source of data. Regardless of the data collection method, the psychologist is only interested in those findings which can be readily verified and duplicated under similar conditions. The psychologist would refer to this requirement as "replication" of data.

It is extremely important that the reader understand from the onset several points about the field of psychology. First, most people who have not been exposed to the field in any formal fashion associate the psychologist almost exclusively with the study and treatment of mental illness and abnormal behavior. While it is true that within the profession the specialty of clinical psychology and to some extent that of counseling psychology do emphasize the study and treatment of the abnormal behavior of man, the remaining specialties are more concerned with the study of "normal" behavior.[1] After all, most of our society probably would be classified under the heading of normal (if we stretch a point here and there); thus, the psychologist who wants to understand human behavior and develop theories and principles about such behavior will usually confine himself to these "normal" people for purposes of data collection.

A second point is that the term *psychology* is not synonymous with "common sense." Often when "common sense" is applied, the conclusions are incorrect because of insufficient information or a confusion of cause and effect. Another term used incorrectly in connection with psychology is *human nature*. People often attempt to explain behavior by attributing it to "human nature." This is merely a verbalism; it does not explain the behavior. Sometimes the term implies that people act in a certain way because of inherited predispositions. The assumption of an inherited predisposition without valid proof does not stand up in the science of psychology.

THE PSYCHOLOGIST

The profession of psychology is rather complex; it consists of many specialties and interests which are in some way similar and in other ways quite dissimilar. In 1966 there were approximately 30,000 psychologists in the United States. These psychologists differed widely in their fields of interest, place and location of employment, etc. Figure 1.1 presents some descriptive information about United States psychologists.

There are more men than women psychologists, although the ratio varies somewhat with particular subfields of psychology. Thus, from Figure 1.1a we see that there are five men for every two women in the areas of clinical and educational-school psychology, while the ratio changes to nine men to every woman in industrial and personnel psychology. Of the total number of psychologists, as shown in Figure 1.1b, by far the largest number are employed by universities and colleges (about

[1] The reader should not be misled into thinking that there exists a sharp and clear dichotomy between the abnormal and the normal in terms of behavior.

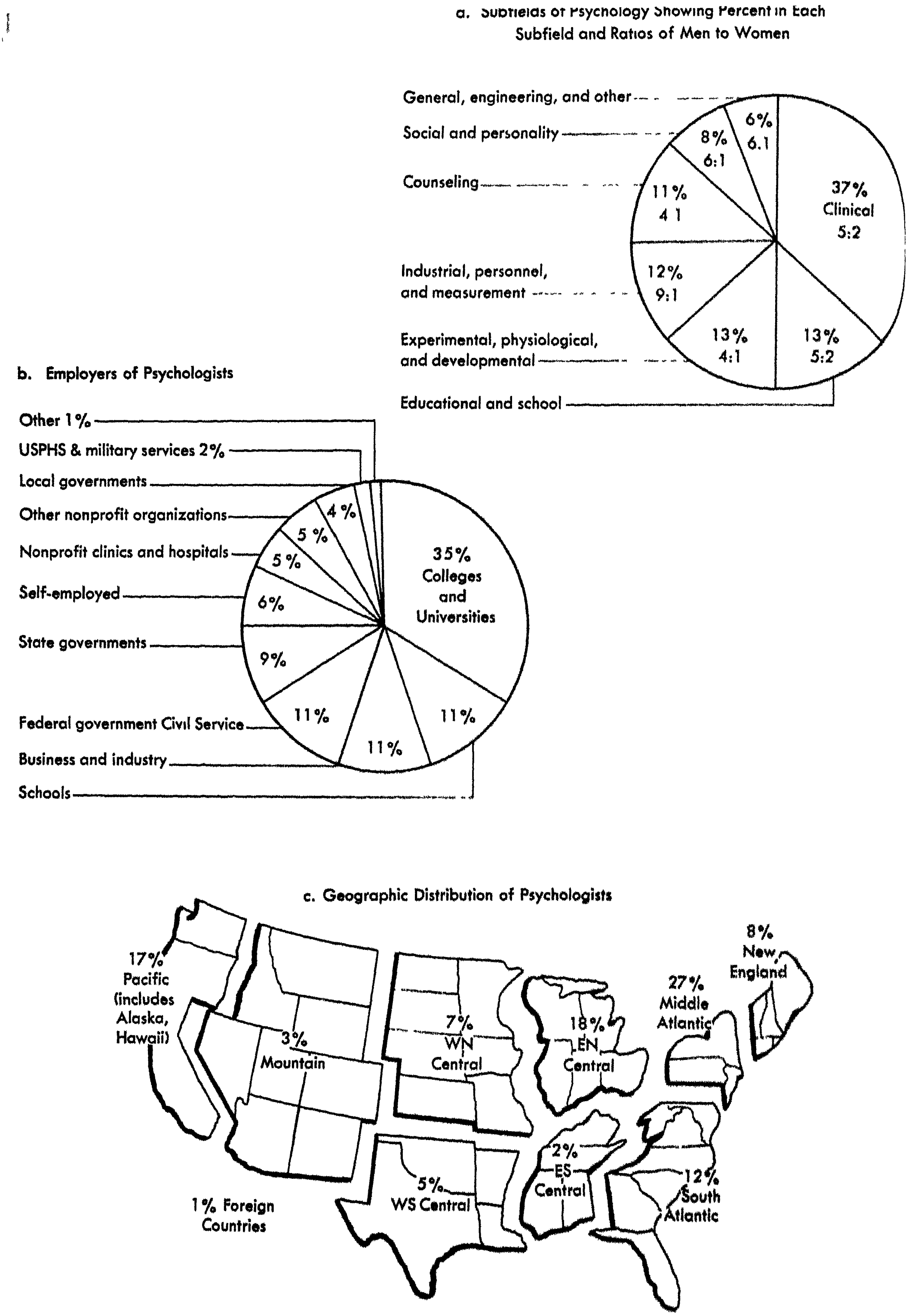


Figure 1.1. Various characteristics of psychologists in the United States. (Adapted from *A career in psychology*. American Psychological Association, Washington, D.C., 1963, 8, 9, 12.)

35 percent), with federal, state, and local agencies of all kinds (excluding schools) employing the second largest number (about 25 percent). Other substantial categories of employment for psychologists are business and industry (about 11 percent) and nonprofit organizations (about 10 percent). Only a relatively small percentage (6 percent) of psychologists are self-employed. Geographically, over one-fourth of all United States psychologists are located in the Middle Atlantic states; the East North Central states (18 percent) and the Pacific states (17 percent) are the other two major geographical locations with substantial numbers of psychologists.

Although the majority of psychologists appear to be located in the coastal states, the majority of the larger training programs are located in the Midwest. Table 1.1 lists the 13 departments with the largest number of Ph.D. graduates in psychology during the academic year 1962–1963. Nine of the thirteen are located in the Midwest, three are Eastern schools, and one is located on the West Coast.

TABLE 1.1 *Doctorates Awarded by the Largest Departments in the Academic Year 1962–1963*

Department	*Doctorates Awarded*
1. Columbia University	45
2. University of Michigan	43
3. Purdue University	40
4. Ohio State University	22
5. University of Minnesota	22
6. New York University	22
7. State University of Iowa	21
8. University of California (Berkeley)	20
9. Indiana University	20
10. Michigan State University	18
11. University of Massachusetts	16
12. Western Reserve University	15
13. University of Illinois	15

SOURCE: Adapted from R. F. Lockman and S. Ross. Survey of graduate education in psychology: Some trends of the last decade. *American Psychologist,* 1964, *19,* 623–628.

Table 1.2 also shows some interesting figures on the relative numbers of psychologists with different specialties who were given degrees the same year (1962–1963). By far the largest percentage of those receiving the Ph.D. degree were those specializing in clinical psychology. Note that industrial Ph.D.'s accounted for only 5.2 percent of the total. With the master's degree quite a different story appears: Those specializing in industrial psychology are the largest single group (almost 25 percent), followed by the specialty areas of educational, experimental, and clinical psychology in that order.

TABLE 1.2 *Types of Graduate Degrees Granted in Psychology in the Academic Year 1962–1963*

	Doctorates		*Masters*	
Area of Specialization	*Number of Degrees*	*Percent of Field*	*Number of Degrees*	*Percent of Field*
Clinical	320	36.4	290	16.1
Experimental	257	29.2	301	16.8
Social and personality	84	9.5	57	3.2
Industrial	46	5.2	446	24.8
Counseling and guidance	43	4.9	157	8.7
Educational	36	4.1	370	20.6
General	28	3.2	23	1.3
Developmental	20	2.3	19	1.1
Psychometrics	19	2.2	35	1.9
School	15	1.7	60	3.3
Engineering and human factors	2	0.2	13	0.7
Unspecified and others	10	1.1	25	1.4
Total:	880		1796	

SOURCE: Adapted from R. F. Lockman and S. Ross. Survey of graduate education in psychology: Some trends of the last decade. *American Psychologist, 19,* 623–628.

INDUSTRIAL PSYCHOLOGY

Industrial psychology draws upon the facts, generalizations, and principles of psychology. It uses the method prescribed in the parent body. Because it applies the techniques of psychology to the industrial scene and the problems confronting it, industrial psychology formulates and modifies procedures to meet the conditions found in business rather than in the laboratory.

Industrial psychology is simply the application or extension of psychological facts and principles to the problems concerning human beings operating within the context of business and industry.

The most important aspect of industrial psychology is its discipline. It clearly recognizes that scientific conclusions must be objective and based upon facts gathered as a result of a defined procedure. It does not treat assumptions or hypotheses as if they were conclusions. Its findings sometimes confirm the obvious and very often are not romantic. Whereas chemists are no longer confused with alchemists, and psychologists are differentiated from physiognomists, the same clear-cut differentiation does not apply to industrial psychologists in relation to their "gold brick" salesmen breathren who sell the impossible with all the authority of ignorance.

The misinformed often believe that all that is necessary to have a knowledge of psychology is to be a human being with common sense and to be a student of human nature. But a psychologist is quite a different person. He has professional training, is aware of the limitations of his knowledge, and above all, has been indoctrinated in a methodology to obtain conclusions based upon data acquired according to certain prescribed scientific methods.

THE INDUSTRIAL PSYCHOLOGIST

The industrial psychologist may be employed by a company. He may be a consultant to a trade association, to retailers or manufacturers. He may also be a university professor who does research in the field either privately as a consultant or as an associate in the university institute. In the future he may be on the staff of a labor union. Regardless of the field in which he works, the industrial psychologist must report research findings accurately. Hence, there is a great need for an unbiased approach if he is to make his full contribution to society.

Although industrial psychologists have been employed predominantly by large companies, this does not mean that small business organizations cannot benefit from the correct application of the principles and generalizations of industrial psychology. As more students of psychology become interested in the industrial field, and as more small businessmen are made aware of the benefits of industrial psychology, greater advantages will accrue to our society. Because small and large business organizations have similar problems, the findings are likely to be applicable to both.

Since many states do not license industrial psychologists—or any psychologists, for that matter—the question arises as to how one becomes an accredited industrial psychologist. Legitimate recognition occurs primarily along educational and professional lines. The American Psychological Association (APA) is the professional body of psychologists; similar associations exist among dentists, physicians, sociologists, etc. The APA has three classes of membership. The *Associate* is a person who meets such minimum standards as either two years of graduate work in psychology at a recognized graduate school or a master's degree in psychology from a recognized graduate school. At the time of application the candidate must be devoting full time to either graduate or professional work in psychology. The minimum standard for election to APA's *Member* status is the receipt of a doctoral degree based in part upon a psychological dissertation and conferred by a graduate school of recognized standing. *Fellows* represent a higher status of membership; they have the Ph.D. degree, five years of acceptable professional experience subsequent to the granting of the degree, and are elected to Fellow status for outstanding scientific achievement in their chosen area of specialization. Life membership, not likely to be of immediate concern to students, is reserved for members of twenty years' standing who have reached sixty-five years of age.

One of the divisions of the American Psychological Association is Division 14, known as the Division of Industrial Psychology. It has three classes of membership, Associate, Member, and Fellow, which require similar standards as membership in the APA except that the experience requirement must be related to industrial experience as a psychologist In 1966 this division had approximately 875 members; the entire APA had about 24,000 members. Not all psychologists belong to the APA, of course. In 1966 there were an estimated 6000 psychologists who met the requirements but did not join.

A further degree of attainment outside of APA is the diploma, which is a certification of specialty. This diploma is awarded by the American Board of Examiners in Professional Psychology. Certification procedures have been set up in clinical, counseling and guidance, and industrial psychology. Standards for diplomate status are the Ph.D., five years of qualifying experience, and the passing of written examinations in the specialized field plus an oral examination after the written. In 1966 there were

about 1900 diplomates; about 180 of these held their diploma in industrial psychology.

Sawyer (1960)[2] reports some interesting facts about the membership of Division 14 taken from a survey of 708 Division members in 1959 (there was a total membership of 716 that year). Of the 81 percent with the Ph.D., over half had received their degrees since 1948, with the breakdown by decades being:

1900–1909	1
1910–1919	6
1920–1929	32
1930–1939	120
1940–1949	160
1950–1958	252

TABLE 1.3 *Schools Granting Degrees in Industrial Psychology 1906–1958*

School	*Number of Graduates*
Purdue University	83
Columbia University	63
Ohio State University	63
New York University	41
University of Chicago	30
University of Minnesota	29
State University of Iowa	26
Western Reserve University	26
University of Michigan	22
University of Pennsylvania	21
Harvard University	18
Northwestern University	17
Pennsylvania State University	16
University of Pittsburgh	16
Stanford University	15
University of Southern California	15
Cornell University	11
Yale University	11
University of California (Los Angeles)	10
University of Wisconsin	10
University of California (Berkeley)	9
66 other schools	156
Total	708

SOURCE: Adapted from J. Sawyer. The industrial psychologist: education and employment. *American Psychologist*, 1960, *19*, 670–673.

[2] Full sources for authors or works mentioned but not footnoted are given in the section "References" which appears at the end of each chapter.

Those universities which Sawyer found to be most productive in graduating Ph.D.'s in industrial psychology are shown in Table 1.3. Sawyer's survey also indicated that the general membership in Division 14 could be categorized into four general employer groups:

Industry	35%
University	26%
Consulting firms	28%
Government	11%

Percentages extremely close to these were found in a later study by Dunnette (1961).

Of course, not all industrial psychologists belong to the APA or to Division 14. Although it is exceedingly difficult to obtain complete figures, it has been estimated that there were about 2500 psychologists employed on a full-time basis in industry in 1963 (Ross and Lockman, 1963) over three times the number belonging to Division 14. This is considerable growth from earlier estimates of 1000 in 1958 (McCollum, 1959) and of 600 estimated by the APA in a 1954 article on the profession (APA, 1954). It is worth noting that these numbers probably represent lower bound estimates of the number of psychologists involved in the activities of business and industry at each of these time periods. Traditionally, many of the psychologists who hold academic positions at various universities spend a portion of their time in a consulting capacity with various business organizations.

SCOPE OF THE INDUSTRIAL PSYCHOLOGIST'S WORK

Perhaps one of the best ways to begin to understand the complex field of industrial psychology is by enumerating the various types of activities which can be considered as part of the general domain of the profession. Actually, even this is not an easy task. Many authors have attempted to set up classifications for describing the activities of the industrial psychologist, with various degrees of success (e.g., see McCormick, 1955; Stagner, 1957; and Taft, 1946). Several have taken the approach that industrial psychology must be "what industrial psychologists do" and have tried to empirically determine exactly what it is that they do. For example, Canter (1948) analyzed the results of a questionnaire returned by 56 respondents in business and industry, 37 consultants, and 10 psychologists in the field of advertising. The results indicated that employers were generally large corporations, but that consultants' clients were both large and small companies. The psychologists in advertising were generally employed by advertising agencies. Table 1.4 presents an analysis of most important work functions of the respondents.

McCollom (1959) actually interviewed 75 psychologists in 20 different cities, having each one describe his own activities and the activities of other psychologists working in industry with whom they were familiar. He found that the activities of these people could be grouped into the following general categories:

1. *Personnel selection:* Selection and assessment of employees and executives, criterion research
2. *Personnel development:* Performance appraisal, attitude measurement, management development, employee counseling

3. *Human engineering.* Equipment and product design
4. *Productivity study:* Activities concerned with worker fatigue, lighting, general job environment
5. *Management.* Activities involving administrative skills
6 *Other:* Accidents and safety, labor relations

TABLE 1.4 *Analysis of Most Important Work Functions, by Groups*

	Percentage of Time
Business and industry group (N 56)	
Duties involving:	
Personnel functions	33
Scientific, research, and developmental functions	25
Policy and management functions	20
Labor relations functions	12
Education and training functions	5
No response	5
	100
Consulting group (N 37)	
Duties involving:	
Personnel functions	25
Policy and management functions	24
General consulting functions	21
Research and development functions	11
Clinical functions	10
Labor relations functions	3
Market research functions	3
No response	3
	100
Advertising group (N 10)	
Duties involving:	
Market research functions	70
Policy and management functions	30
	100

SOURCE: R. R. Canter. Psychologists in industry. *Personnel Psychology,* 1948, *1,* 145–161.

Perhaps the most authoritative source on the functions and activities which comprise industrial psychology is Division 14 itself, since it exists as the official organization of all psychologists concerned with the application of psychological knowledge to the world of industry. In a 1959 report entitled *The Psychologist in Industry,* the Division cited seven major areas which comprised the content area of industrial psychology. These were: (1) selection and testing, (2) management development, (3) counseling, (4) employee motivation, (5) human engineering, (6) marketing

research, and (7) public relations research. These were very similar to the functions outlined by Taft (1946) who listed job analysis; motion studies, salaries and wages, selection of new employees; transfers, promotions, and terminations; training; problem employees; employee rating; industrial hygiene; morale and research. In summary, it would appear that the psychologist employed by an industrial organization is likely to find himself required to tackle nearly any problem involving people. This is made quite apparent by the Division 14 report, which lists over 150 more specific kinds of problems with which industrial psychologists concern themselves.

An example of the varied nature of the work is given in an article by Bills (1934) describing a typical workday of an industrial psychologist. The day started with a conference at 9 A.M. which resulted from an interview held the preceding afternoon. The conference was to decide on the placement of an employee for a two-month

TABLE 1.5 *Representative Firms Employing Diplomates in Industrial Psychology and Diplomates' Titles*

Firm Name	*Title*
American Home Products Corp.	Director of Personnel
Atlantic Refining Co.	Research Assistant
B. F. Goodrich Co.	Coordinator of Training
Carbide & Carbon Chemical Co.	Asst. Director, General Industrial Selection Department
Chrysler Corp.	Educational Supervisor
Columbia Broadcasting System, Inc.	President
Commonwealth Life Insurance Co.	Director of Research
Continental Oil Co.	Director of Advertising
D. E. McNicol Pottery Co.	Vice-President and General Manager
Detroit Edison Co.	Industrial Psychologist
E. I. du Pont de Nemours & Co.	Manager, Personnel Research Section, Pres. Div
Fieldcrest Mills	Department of Personnel Research & Training
General Motors Corp.	Chairman, Personnel Evaluation Services
General Shoe Corp.	Director Supervisory Training
Gulf Oil Co.	Head, Training Unit
Harwood Manufacturing Co.	President
International Business Machine Corp.	Coordinator, Educational Research
Knox Reeves Advertising, Inc.	Vice-President, Marketing Director
Life Insurance Agency Management Association	Research Associate
Marshall Field Co.	Vice-President
Needham, Louis and Brorby	Vice-President
Procter & Gamble	Head, Research Department
Prudential Insurance Co.	Asst. General Manager, Director of Personnel Research
S. C. Johnson & Son, Inc.	Personnel Director
Standard Oil Co.	Advisor, Employee Relations Research
United States Steel Corp.	Manager, Personnel Department
Wm. Esty Co.	Executive Vice-President

period. This employee had been diagnosed as mentally ill, and the company was attempting therapy by a psychiatrist rather than resorting to immediate dismissal. The day's second problem for the psychologist was an attempt to predict whether a person's ability to punch Hollerith cards could be ascertained in three weeks. The next problem concerned the transfer of two employees to fill two vacancies. Problems concerning salary levels and rating scales occupied the remaining portion of the morning. Immediately after lunch the psychologist had to meet with the Employee's Loan Fund Committee; the particular problem facing them concerned a misapplication of some funds. The psychologist next talked with a cleaning woman who felt that the supervisor was not giving her a fair deal. Another employee who had been married for six months asked the psychologist for advice on how to inform her disapproving parents.

Actual test development and research thus played only a small role in this particular day's work. The psychologist who is on the staff of an organization is likely to find that the specific minor problems which arise each day may interfere with his major work. He must therefore be flexible and prepared to handle a number of tasks and projects concurrently.

Not only are the duties of the industrial psychologist varied, but his title ranges from *President* to *Industrial Psychologist* to *Educational Supervisor*. Table 1.5 lists business and industrial firms that employ diplomates in industrial psychology and their titles. The firms selected for inclusion in this table are merely representative. The table is intended to give an idea of the range of titles and not to be a complete listing of companies employing industrial psychologists.

Table 1.5 reveals that few industrial psychologists are assigned the title of the profession. The clear picture is that the psychologist is assigned a title related to his job function or duty.

In addition to the diplomates employed in business and industry, psychologists work for many other companies, including the following. Again, no attempt is made to furnish a complete list but, rather, to give an idea of the wide range of companies employing psychologists.

Aetna Life Insurance Co.
American Can Co.
American Viscose Corp.
Armco Steel Corp.
Caterpillar Tractor Co.
Continental Can Co.
Corning Glass Works
Creole Petroleum Corp.
Eli Lilly & Co.
Esso Standard Oil
Ford Motor Co.
General Motors Corp.
Hughes Aircraft
International Harvester Co.
Kimberly-Clark Corp.
McCann Erickson, Inc.
Metropolitan Life Insurance Co.
Midland Cooperative Wholesale
Minneapolis Honeywell Regulator Co.
Minnesota Mining & Manufacturing Co.
Pittsburgh Plate Glass Co.
Rand Corp.
RCA: Victor Division
State Farm Insurance Co.
The Chesapeake & Ohio Railway Co.
The Dayton Co.
The Prudential Insurance Co. of America
Washington Gas Light Co.
Young & Rubicam

CONSULTING ORGANIZATIONS

An industrial psychologist, in addition to being directly employed by a business organization, is also found in many consulting organizations. Today consulting on problems of a psychological nature is "big business." Most psychological consulting firms may be classified into one of two possible categories: Those who emphasize the area of psychological testing and those who emphasize the nontesting aspects of industrial psychology. Admittedly the separation is generally not a distinct one, particularly with the very large consulting firms who tend to handle both categories. However, the emphasis usually still remains visible. Following are some brief descriptions of several of the most prominent consulting firms typical of each of the above categories.

CONSULTING FIRMS: TESTING

The oldest and largest organization of this type is the Psychological Corporation which was organized in 1921 by a group of psychologists. The Psychological Corporation is organized into the following divisions: Industrial, Test, Personnel and Market Research, Professional Examinations, and the Experimental Laboratory.

The Industrial Division installs testing programs for companies, improves training programs for management personnel, and trains supervisors in various techniques and principles in human relations. The Test Division publishes and sells psychological tests to schools, government agencies, and industrial firms. The Personnel and Market Research Division does research on consumer motivation and behavior, public relations, and media; it also carries on activity in personnel selection attitudes and job evaluation. The Professional Examinations Division has evaluated several hundred thousand nursing school applicants and currently also conducts entrance examination programs for schools of veterinary medicine and pharmacy. The final division, the Experimental Laboratory, undertakes research projects in applied psychology, such as the development of test scoring machines, research in logical abilities, etc. The Psychological Corporation employs about 25 Ph.D.'s in its various divisions.

A second major consulting firm which emphasizes psychological testing is Science Research Associates (SRA). This corporation specializes in developing material and services for education, industry, and government based upon research in the behavioral sciences. It offers a variety of publications and services for instruction and training, educational and vocational guidance, personnel selection and placement, evaluation of knowledge and performance, and the processing of specialized data.

The third of the "big three" in testing is the Educational Testing Service (ETS), a nonprofit organization founded in 1947 through the combined efforts of the American Council on Education, the Carnegie Foundation, and the College Entrance Examination Board. Its major purposes are to develop tests for schools, colleges, and the psychological profession; to assist the users of tests in methods and procedures; and to do research in psychological testing. In 1966 over 70 psychologists were listed by the APA as being affiliated with ETS (APA Directory, 1966). In the fiscal year ending June 30, 1966, ETS had a gross income of over $23 million —ample testimony to the importance of testing in our current culture.

CONSULTING FIRMS: GENERAL

The firm of Richardson, Bellows, Henry and Company, Inc. was organized shortly after World War II. Its major office is in New York City, but it has branch offices in other major cities. This company primarily deals with employee attitude measurement, executive appraisal, the development and improvement of training programs, manuals for client companies, and research in test development and employee evaluation.

Dunlap and Associates, Inc. is a consulting firm organized in 1948 to provide research and consulting services in the fields of executive evaluation, marketing research, human engineering, and training. Its staff now numbers approximately 200, with about one-fourth of the professional staff possessing a Ph.D. In its 1964 Annual Report, it showed a gross income of nearly $4 million (see Figure 1.2).

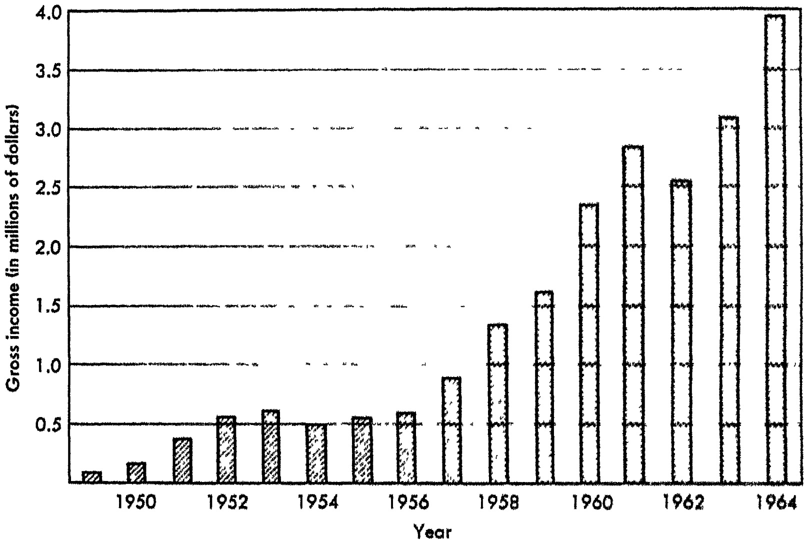


Figure 1.2. Gross income of Dunlap and Associates, Inc. from 1949 to 1964. (Adapted from the 1964 Annual Report, Dunlap and Associates, Inc., Stamford, Conn.)

Rohrer, Hibler, and Replogle have offices in six cities and primarily provide psychological services to business leaders. Their approach is first to obtain a psychological evaluation of the "problem" executive and then to conduct a series of individual conferences with him designed to change his behavior into more desirable patterns of action (Flory and Janney, 1946). This company was founded in 1927 and has five senior partners, all of whom have a Ph D.

The American Institutes for Research in the Behavioral Sciences (AIR) is a nonprofit research organization which carries out a variety of research programs in human engineering, training, organizational behavior, and education. Its major offices are in Pittsburgh, with subsidiary offices in Washington, Los Angeles, and Palo Alto. During the 1962–1964 period AIR's average number of full-time employees was around 230 (AIR Annual Report, 1964).

RESEARCH CENTERS AT UNIVERSITIES

The practical aspect of solving problems by using research techniques was intensely demonstrated during World War II. The armed forces did not have the personnel or the unlimited facilities to do all the work that was needed. Accordingly, many universities established research centers to handle problems related to industrial psychology and other areas of psychology. This situation has continued, and now there are many such centers handling on a contract basis research needed by many government agencies.

Although research in personnel problems and human engineering is of primary interest for our subject matter, additional work is done in the fields of physiological and social psychology, and in such topics of general psychology as learning, retention, and the higher cognitive processes.

Research contracts are awarded by the Department of Defense as well as other branches of the government such as the Atomic Energy Commission, the Department of the Interior, the Veterans' Administration, and the National Aeronautics and Space Administration.

Specific problems related to job analyses and specifications; selection, classification, and criteria research; training; psychometrics; and human engineering are solved by contract research. Examples of nonsecurity contracts are as follows: billet analyses for guided missiles personnel; a study of the relationships between Navy billets and civilian occupations; submarine personnel selection; research on supervisor selection; radar mechanics functional knowledge test battery; effectiveness in technical training; speed factors in tests and in criteria; causation of accidents; the worker as a factor in equipment design; display and signal pattern discrimination; and attention value of warning signals. There are many, many others.

Possibly the largest university research center in psychology is at the University of Michigan. It is known as the Institute for Social Research and has two main divisions, the Survey Research Center and the Research Center for Group Dynamics. Although the staff of highly competent professionals is primarily concerned with social psychology, the work clearly shows the intimate interrelationship of social and industrial psychology.

The Survey Research Center is more concerned with problem-oriented than with discipline-oriented research. This means it promotes interdisciplinary research combining the efforts of psychologists, economists, anthropologists, sociologists, and political scientists. Its research undertakings are sponsored by government agencies, private business, and research foundations. An example of one of its major concerns is its work in the field of psychological economics and consumers' economic attitudes. On the other hand, the Research Center for Group Dynamics is primarily concerned with the phenomena of group behavior and with an attempt to derive the principles of group dynamics which lead to group formation, change, or dissolution. It has worked in the areas of group productivity, communication, and intergroup relations.

These two centers together with the parent body employ over 500 full- and part-time personnel, including a permanent part-time force of about 250 field interviewers residing in communities across the nation. The organization chart of the Institute (Figure 1.3) depicts its wide range activities. The Institute has published

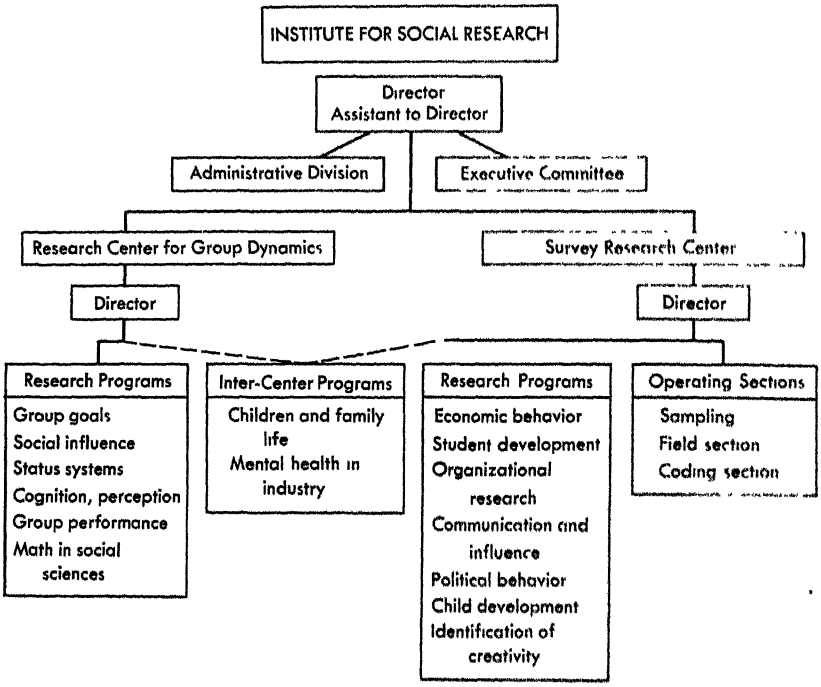


Figure 1.3. Organizational chart describing the structure of the University of Michigan's Institute for Social Research in 1963.

well over 2000 articles, reports, and books as a result of its research activities (Seashore, 1962).

Another notable but less clearly structured interdisciplinary research center is the joint research composite at Ohio State University. This composite is made up of three distinct suborganizations, the Bureau of Business Research, the Behavioral Sciences Laboratory, and the Administrative Services Division.

CENTERS OF LABOR AND INDUSTRIAL RELATIONS

A number of universities have well-established centers or institutes related to the problem of labor and industrial relations. The Industrial Relations Center of the University of Minnesota is concerned with a program of training and research to improve labor-management relations and to facilitate cooperation between the two. The Center recognizes that such work cannot be solely the province of any one department; thus it actually draws from eight different departments, one of which is psychology.

The Institute of Industrial Relations of the University of California has two divisions, one at Berkeley and the other at Los Angeles. It pursues three major lines of endeavor—research, instruction, and community relations—and runs many conferences and lecture series as well as resident institutes for labor and for management. It also accepts the importance of the interdisciplinary approach.

Other outstanding programs of this nature are located at Cornell University (The New York State School of Industrial and Labor Relations) and at the University of Illinois (Institute of Labor and Industrial Relations).

HUMAN ENGINEERING LABORATORIES

Several university psychology departments have research laboratories in which the emphasis has been directed toward research on human engineering topics. Probably the best known of these is the Laboratory of Aviation Psychology (recently renamed the Human Performance Center) at Ohio State University. In addition, the University of Illinois's Aviation Psychology Laboratory and the University of Michigan's Engineering Psychology Laboratory are equally well-established centers of human engineering research.

OTHER RESEARCH ORGANIZATIONS

There are two other important types of organizations in which industrial psychologists are employed: nonprofit organizations sponsored by industry, and various governmental agencies which specialize in research problems relevant to industrial psychology.

NONPROFIT ORGANIZATIONS SPONSORED BY INDUSTRY

A number of nonprofit research and consulting organizations have been sponsored and funded by companies in a common industry to do research for the entire group of member companies. A typical example is the Life Insurance Agency Management Association (LIAMA) formed in 1922. It now has over 360 full-member participating companies and 118 associated members scattered all over the world. It is organized into such divisions as company relations, research, actuarial, institutional relations, and administrative. The three major parts of the research division are Human Resources Research, Financial Management and Functional Cost Research, and Market and Survey Research. Typical projects include studies of agent and managerial selection procedures, training of agents, and the measurement of public attitudes toward life insurance.

The National Industrial Conference Board was founded in 1916 as an independent and nonprofit institution for business fact-finding through research. It is supported by more than 3000 subscribing associates from business organizations, trade associations, labor unions, libraries, and colleges. The research program is carried on by four major divisions. The Division of Personnel Administration is of greatest interest to the industrial psychologist. Its reports serve as useful guides to those who want to learn from case studies as well as have a manual type of reference.

Other examples of such organizations are the Committee on Highway Safety Research and the W. E. UpJohn Institute.

GOVERNMENT RESEARCH AGENCIES

As was pointed out earlier in the chapter, many industrial psychologists are employed by the United States Government. Examples of government agencies which have traditionally been involved in psychological research are listed below:

U.S. Army Personnel Research Office
Washington, D.C.

U.S. Air Force Personnel Research Laboratory
San Antonio, Texas

U.S. Navy Aviation Psychology Laboratory
Pensacola, Florida

U.S. Navy Electronics Laboratory
San Diego, California

U.S. Air Force Behavioral Sciences Laboratory
Dayton, Ohio

U.S. Navy Submarine Base
New London, Conn.

Two of the largest single agency employers of psychologists are nongovernmental research and development agencies. However, both of these agencies work primarily for governmental agencies. HumRRO, the Human Resources Research Office of George Washington University, was established in 1951 to apply scientific methods to the solution of army problems in training, motivation, and leadership. The System Development Corporation (SDC) was later established to perform a somewhat similar function for the Air Force. As of 1964 HumRRO employed approximately 75 research psychologists and SDC employed nearly 150.

The organizations mentioned do not exhaust the list of industries, consulting organizations, centers at universities, cooperative research centers supported by industry, or government agencies. They have been selected merely as examples, but of course all pass rigid test of authenticity and compliance with stated purpose. Others too numerous to mention also meet such standards.

Examples have been cited to impress the reader with the realness of research. Each organization has been described all too briefly, but the bibliographic references can furnish the necessary additional facts.

HISTORY AND DEVELOPMENT OF THE FIELD

It is exceedingly difficult to apply a date to the founding of any discipline. However, industrial psychology may have gotten its start on December 20, 1901. That was the evening that Dr. Walter Dill Scott, a psychologist at Northwestern University, gave an address discussing the potential application of psychological principles to the field of advertising (Ferguson, 1962). This was followed by a series

of 12 magazine articles which were subsequently combined in 1903 into a book entitled *The Theory of Advertising* (Scott, 1903)—undoubtedly the first book which involved the application of psychology to the world of business. Scott, who later became president of Northwestern University, followed up this initial book with several others in the next few years on the same general topic (Scott, 1908a, 1908b). He also published several books in which he tried to bring psychology to bear upon the broader domain of business per se, rather than just advertising (Scott, 1911a, 1911b). In fact, Dr. Scott's accomplishments as the first industrial psychologist are far too numerous to detail here, the reader is referred to more appropriate sources (Ferguson, 1962).

In spite of Scott's prominent activity and many published works during the years 1901 to 1913, it is interesting that he has often been ignored in historical accounts of the field of industrial psychology. This is particularly so in past years. Today he seems to be gradually regaining his rightful place as the man who really "started it all."

Where Scott has not been credited with being the first industrial psychologist the honor has usually been awarded to Hugo Münsterberg. In 1913 he published his text *Psychology and Industrial Efficiency* (Münsterberg, 1913). Münsterberg as an author tended to be somewhat less aggressive and flamboyant than Scott in his approach to applying psychology to business. Thus in his book the differences of approach between the pure and the applied science are carefully reflected in the first few chapters. Münsterberg writes cautiously and defensively of his attempts to establish an applied field of psychology as a necessary counterpart of the pure field. Münsterberg's book, which has served as a model for the development of industrial psychology, includes such topics as learning, adjustment to physical conditions, economy of movement, monotony, fatigue, and buying and selling.

During World War I psychologists were quite active in the war effort, developing group tests for army recruits and aiding in the development of procedures for the selection of officer personnel. In fact, many of the major postwar developmental areas of industrial psychology such as group testing, trade testing, rating scales, and the personality inventory had their roots in the activities of psychologists in the World War I war effort.

In 1917 the *Journal of Applied Psychology* made its appearance; the importance of this periodical in the expansion of the field will be apparent from the number of references throughout this book to articles which have appeared in it. At about the same time, colleges began to introduce courses in applied psychology; and as the subject has developed, the trend has been to offer courses in specific fields of applied psychology such as industrial psychology, personnel psychology, vocational psychology, and advertising psychology.

In 1919 Scott and several others founded the Scott Company of Philadelphia. This was the first psychological consulting firm ever organized and was oriented toward personnel problems in industry. This was followed two years later by the previously mentioned Psychological Corporation—the oldest firm of this type still active in the United States.

During the post-World War I years industry first began to show an interest in the discipline of industrial psychology. Certain firms such as Procter & Gamble, the Philadelphia Company, and the Hawthorne Plant of Western Electric all formed their own personnel research programs. In fact, it was at the Hawthorne Western Electric plant that the famous "Hawthorne" studies were begun in 1927 (Roethlis-

berger and Dickson, 1939). These studies, which lasted into the middle 1930s and which are described in Chapter 10, are considered by many, the present authors included, to be probably the most significant series of research studies ever carried out in industry in terms of the impact they had on the growth and development of industrial psychology. They provided the foundation and impetus for the expansion of industrial psychology beyond the realm of selection, placement, and working conditions into the study of motivation, morale, and human relations.

The Depression itself had considerable effect on the development of industrial psychology. While it may have slowed growth in some directions, it nevertheless opened many additional areas for survey. After the Depression the importance of employee attitudes began to be recognized; consequently much development since that time has been in this area.

Without much question World War II was a major factor in the growth of psychology in industry. Although the American Association for Applied Psychology was formed in 1937 as the official organization of industrial psychology (it later became Division 14 of the APA), it was the huge psychological contribution to the war effort which proved to so many people that applied psychology had important and practical contributions to offer. These contributions were far too numerous to document in any detail here. As in World War I, great emphasis was placed upon the development of tests for selecting and classifying recruits. Also developed were selection programs for officers, various training programs of specialized types, and job analysis and performance evaluation techniques.

Personnel Psychology, another major journal of applied research, first appeared in 1948. It publishes the results of factual psychological studies in such fields as training, job analysis, selection, evaluation, motivation and morale, work conditions, and equipment design. Its articles are intended for interested and informed management and yet at the same time they are designed to meet the technician's requirement of accurate and complete reporting.

Another notable development in applied psychology since World War II was the establishment of other separate divisions of the APA devoted to various aspects of the field: the Division of Military Psychology (Division 19); the Society of Engineering Psychologists, a division of the APA (Division 21); and the Division of Consumer Behavior (Division 23). While the percentage of joint memberships in these divisions and Division 14 is high (i.e., many psychologists belong to two or more divisions), the fact that they have been formed is ample testimony to the need of the psychologist to recognize the ever-increasing complexity and specialization of interest in the web of activities which define applied psychology.

Several other major organizations have been created since World War II to represent various interest subgroups in applied psychology. The two most notable are the Human Factors Society and the Ergonomics Society. The former is an American group for applied psychologists with interest in human engineering problems. It publishes its own journal, *Human Factors*. The latter group is the British counterpart of the Human Factors Society, although it was the first to be organized. Its journal is entitled *Ergonomics*. *Ergonomics* and *Occupational Psychology* are the two major British publications devoted to applied psychological research. Recently (in 1966) another American journal, *Organizational Behavior and Human Performance*, started publication. The purpose of this journal is to publish articles devoted to the development of theory in applied psychology.

MAJOR PROBLEMS OF INDUSTRIAL PSYCHOLOGY

Before proceeding to the methods and content of industrial psychology, it might be best to mention certain major problems which the profession has to face in its future growth and development.

THE CONSULTANT AND THE STAFF PSYCHOLOGIST

As was pointed out earlier, the industrial psychologist is likely to obtain his livelihood through one of three major sources of employment. He is either a consultant, an employee of a company or the government, or a university teacher. Very often he combines two of the three roles, but whether he does or not depends upon his interests, opportunities, degree of identification, and tempo.

A psychologist directly employed full time by a company or by a government agency is often referred to as a "staff" psychologist. Generally speaking, the duties and tasks of the consultant and the staff psychologist overlap. There is no clear-cut difference insofar as type of assignment is concerned. The major difference is that the consultant may be concurrently working for a number of clients or employers, whereas the staff psychologist fills a more specific role in the organization chart for a single employer.

Although a schism between the staff psychologist and the consultant is undesirable if the profession is to be advanced in industry, the answers given in Canter's study (1948) to the question "What do you think of consulting firms as the best solution to industrial psychological problems?" pose a serious future problem. One-half of the staff psychologist group was unfavorable toward such firms; the consulting group was generally favorable. This situation demands attention and should be cleared up.

A note of optimism is reflected toward the field in general since 80 percent of the respondents reported that executives were becoming more "psychological minded." A further indication of the increased acceptance of the psychologist by leaders in industry comes from a 1962 survey conducted by Feinberg and Lefkowitz (1962). They administered a questionnaire to 89 executives who were attending a seminar sponsored by the American Management Association. When asked whether they would be interested in hiring an industrial psychologist, over two-thirds replied favorably. These "yes" respondents felt the industrial psychologist could be of greatest benefit in the areas of employee motivation, employee selection and training, executive selection and training, human engineering, consumer research, production efficiency, and accident control.

COMMUNICATION

One of the difficulties of any profession is that its language and techniques sometimes become so involved that the outsider is really left out. If industrial psychology is to gain an important place in industry, psychologists must learn to talk and write in a fashion that is clearly understandable to others who are equally interested in the mutual problems and who sometimes have an even greater stake in a solution. Not only must the industrial psychologist learn to communicate adequately with the

nonpsychologist, but even the problem of communication within the field itself is becoming a problem. The ever-increasing complexity of industrial psychology and the specialization of interest of the psychologists working on different problems in different settings has created many barriers to the flow and dissemination of knowledge among researchers and practitioners. While such problems may be the inevitable corollary of a dynamic discipline, the authors feel that the communication problem is one of the most critical in industrial psychology today.

RESISTANCE TO CHANGE

Research findings as well as research itself can ordinarily be expected to meet with resistance on the part of employees and, in many instances, employers. The successful practitioner of industrial psychology must be immediately and forever aware of this phenomenon. It would be purely academic if one anticipated that industry is waiting with open arms to apply the knowledge of industrial psychology.

Attempts at change, no matter how well-intentioned, produce threats and will be resisted. This resistance may take the form of hostility and aggression against the change itself or against the administrator of the projected change. Often the employee imagines the nature of the change well in advance of the possibility of a change. The unreality of the imagination only makes the resistance stronger. When changes are associated with speedups or layoffs, the resistance to any contemplated change is even more intense. It is not enough to state that no detrimental action to the employee's welfare is contemplated. The claim must be proved. Anything that is not clearly understood can be an insecurity producing factor. Change often upsets established pattern. People are not easily corrected, nor are they able to give up habits freely. Research often intends to change behavior that has become routine and thus can be expected to be resisted.

Resistance comes not only from the employee but from all levels of management and the employer. The naive employer often wants research to prove his point or position. Such a guarantee is not possible: The conclusions of research depend upon the data and cannot be established by manipulation of data to conform to a preestablished outcome.

All, however, is not hopeless provided at least four fundamentals are recognized. First, the reasons for the contemplated change should be clearly explained. Second, those who will be involved in the change should have ample opportunity for participation in the implementation of that change. Third, change should be a two-way affair rather than an attempt to force all to agree to a one-sided decision. Fourth, the administrator of the change should recognize at all times that change is a real, imagined, or potential threat, and that he must do all he can to eliminate or reduce the possible threat regardless of the form it assumes.

CHAPTERS TO FOLLOW

Each of the chapters in this book will discuss one segment of the various duties of the industrial psychologist. There are, of course, many sequences in which the chapter could be presented and many topics which could be excluded or included.

The material is presented, however, according to the way the authors perceive their subject matter. It will be well to remember that each chapter is related closely to all the other chapters, and that chapter divisions are merely educationally desirable tools—they are not the realities. The individual in a work situation, be he employer or employee, has attitudes, satisfactions, motives, and abilities, and many varieties of feelings and knowledge are always in force at the same time.

REFERENCES

American Institutes for Research in the Behavioral Sciences *Annual report* Pittsburgh, 1964

American Psychological Association. *Directory* Washington, D C., 1966

American Psychological Association. Psychology and its relation with other professions Washington, D.C., 1954.

American Psychological Association, Division of Industrial Psychology. *The psychologist in industry.* Washington, D.C., 1959.

Bills, M. A. A day in the life of an industrial psychologist. *Personnel Service Bulletin,* 1934, no. 5

Canter, R. R. Psychologists in industry *Personnel Psychology,* 1948, *1,* 145–161.

Dunlap and Associates, Inc. *Annual report.* Stamford, Conn., 1964.

Dunnette, M. D. Incomes of industrial psychologists *American Psychologist,* 1961, *16,* 533–536.

Educational Testing Service. *Annual report* Princeton, N.J , 1962–1963, 1965–1966.

Feinberg, M. R., and J. Lefkowitz Image of industrial psychology among corporation executives. *American Psychologist,* 1962, *17,* 109–111.

Ferguson, L. W. *The heritage of industrial psychology.* Finlay Press, Hartford Conn., 1962.

Flory, C. D., and E. T. Janney. Psychological services to business leaders. *Journal of Consulting Psychology,* 1946, *10,* 115–119

McCollum, I. N. Psychologists in industry in the United States. *American Psychologist,* 1959, *14,* 704–708.

McCormick, E. J. Role of the psychologist in industry. *Journal of Personnel and Administrative Industrial Relations,* 1955, 2, 23–31.

Münsterberg, H. *Psychology and industrial efficiency.* Houghton Mifflin, Boston, 1913.

Roethlisberger, F. J., and W. J. Dickson. *Management and the worker* Harvard University Press, Cambridge, Mass., 1939.

Ross, S., and R. F. Lockman. A career in psychology. American Psychological Association, Washington, D.C., 1963.

Sawyer, J. The industrial psychologist. education and employment. *American Psychologist,* 1960, *15,* 670–673.

Scott, W. D. *The theory of advertising.* Small, Maynard and Company, Boston, 1903.

Scott, W. D. *The psychology of advertising* Small, Maynard and Company, Boston, 1908(a).

Scott, W. D. *The psychology of advertising in theory and practice.* Small, Maynard and Company, Boston, 1908(b).

Scott, W. D. *Increasing human efficiency in business.* Macmillan, New York, 1911(a).

Scott, W. D. *Influencing men in business.* Ronald, New York, 1911(b).

Seashore, S. E. A description and analysis of institutional context for social research. Paper presented at the meeting of the American Sociological Association, Washington, D.C., August, 1962.

Stagner, R. Some problems in contemporary industrial psychology. *Bulletin, Menninger Clinic,* 1957, *21,* 238–247.

Taft, R. The staff psychologist in industry. *American Psychologist,* 1964, *1,* 55–61.

A STATISTICAL RATIONALE FOR SELECTION AND PLACEMENT

2

The most traditional activity of the psychologist in industry has been employee selection and placement. The challenging task of trying to distinguish "good" workers from "poor" workers with the aid of tests and other selection devices may be said to coincide with the early formulations of industrial psychology. These problems attracted such pioneer industrial psychologists as Hugo Münsterberg, Walter Dill Scott, and Walter V. Bingham into the world of industry in the early 1900s. Münsterberg used psychological tests prior to 1910 in connection with various selection problems for the Boston Elevated Railway Company. Scott and Bingham worked on the problem of selecting salesmen in the period 1915–1917 while they were colleagues at the Carnegie Institute of Technology.

MANPOWER PLANNING

People represent a [illegible] important resource to a business or industrial firm. Thus the terms *selection* and *placement* designate separate phases in the ever-important area of manpower planning. The trends toward automation and computerization are making the selection and placement of people more important rather than less important. Despite any or all technological advances, profits resulting from a company's efficient operation require the expeditious use of manpower via correct selection and placement.

Selection, as the name implies, involves picking for hire a subset of workers from the total set (population) of workers available for hire at any given moment in time. Efficient selection is therefore a nonrandom process, insofar as those selected have been chosen on the assumption that they are more apt to make "better" employees than those who have been rejected. The task of the industrial psychologist is to make certain that the assumption is indeed a valid one as a result of using objective and scientific procedures and instruments rather than subjective and biased judgments. An equally critical problem is that of assigning the new hires to the jobs available. Placement, in its simplest form, is "given the hiring of N men to fill N different job vacancies, which man should be put on which job?" As might

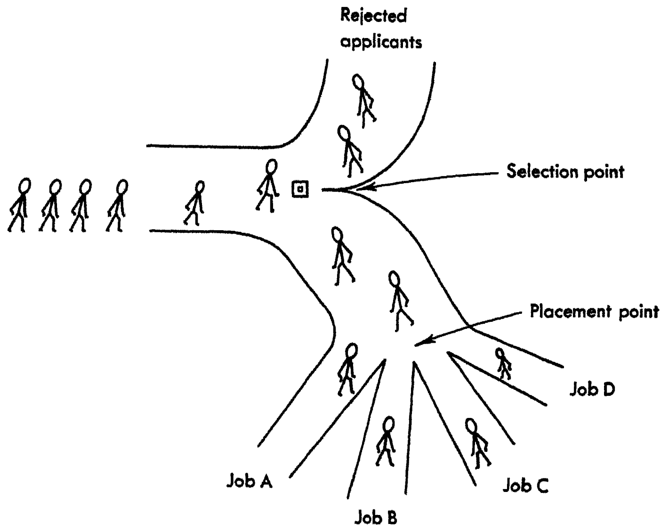


Figure 2.1. Highly simplified diagram of the manpower allocation process.

be suspected, in actual practice the processes of selecting and placing employees often merge into a single process. Figure 2.1 shows in highly simplified form the general process of manpower allocation.

INDIVIDUAL DIFFERENCES

That people differ is self-evident. Measuring any characteristic of people with a measuring instrument will result in a *distribution* of scores—some people will have high scores, others low scores, etc. As an area of psychology, the study of individual differences has provided the impetus for the entire psychological testing movement. Psychological tests measure the extent of differences among people and have been used to obtain estimates for a wide variety of human characteristics.

The existence of individual differences provides the basis for the selection and placement process. Any group of workers, irrespective of occupation, will vary considerably in terms of their relative work efficiency and performance. For example, many years ago C. L. Hull reported some ratios of efficiency between the most efficient and the least efficient workers in a variety of different occupations (Table 2.1). Note that in the case of one occupation (spoon polishers) the best worker was over five times as efficient as the poorest worker.

It is not surprising that employers show considerable interest in wanting to develop ways of selecting those workers who will be more efficient and thus reducing the possibility of hiring those who may be relatively inefficient. This is particularly true for those employers who go to considerable expense training new employees before determining whether they will be successful; and it is very important when union contracts prevent terminating employees for any reason after a brief trial period (often no longer than six weeks).

Another example of how widely workers at the same plant and even at approximately the same job level can vary on a single trait is shown in Table 2.2. The

TABLE 2.1 *Ratio of Least Efficient to Most Efficient Individual in a Variety of Gainful Occupations*

Occupation	*Criterion*	*Ratio of Poorest to Best Worker*
Heel trimmers (shoes)	Number of hours per day	1:1.4
Loom operators (silk)	Percent of time loom is kept in operation	1:1 5
Hosiery maters	Hourly piecework earnings	1:1.9
Loom operators (fancy cotton)	Earnings	1:2
Knitting machine operators	Pounds of woman's hose produced per hour	1:2.2
Elementary school teachers	Ratings of superiors	1:2.5
Spoon polishers	Time per 36 spoons	1:5.1

SOURCE: C. L. Hull. *Aptitude testing*. Harcourt, Brace & World, New York, 1928.

table presents the scores of nearly 2000 workers in single plant on the *Wonderlic Personnel Test,* a short paper-and-pencil test of general mental alertness or intelligence developed specifically for use in industry. For a person unfamiliar with the diversity of talent likely to be found at any job in any industry the range of scores may seem hard to believe—the ratio of ability in terms of IQ equivalent is nearly 2:1 between the top and bottom worker.

The same data are shown plotted graphically in the form of a frequency polygon in Figure 2.2. This illustrates that the distribution of many characteristics of people,

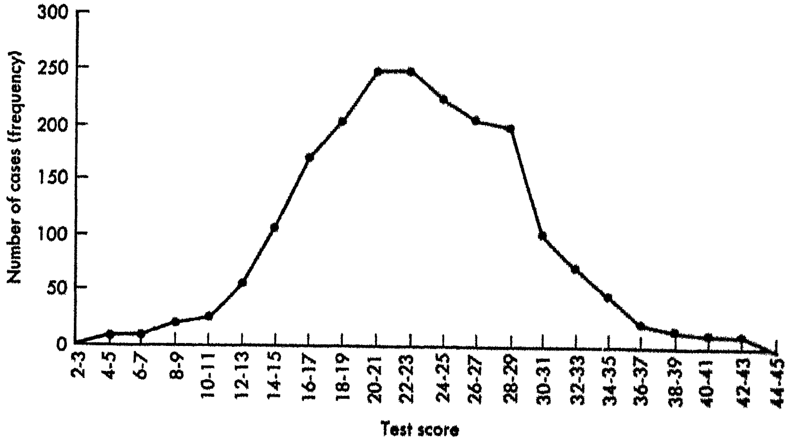


Figure 2.2. Frequency distribution showing the number of employees obtaining various scores on the Wonderlic test in a manufacturing plant. (Adapted from W. J. McNamara. Retraining of industrial personnel. *Personnel Psychology,* 1963, *16,* 233–247.)

TABLE 2.2 *Distribution of Scores on Wonderlic Test for a Random Sample of 1940 Employees in a Manufacturing Plant*

Raw Scores	*IQ Equivalent*	*Percentile*	*Number of Cases*
42–43	131	100.0	4
40–41	129	99.8	4
38–39	127	99.6	8
36–37	124	99.2	15
34–35	123	98.4	43
32–33	119	96.2	69
30–31	116	92.6	100
28–29	113	87.5	197
26–27	110	77.3	203
24–25	107	66.8	227
22–23	104	55.2	249
20–21	101	42.3	246
18–19	97	29.6	194
16–17	93	19.6	164
14–15	89	11.2	105
12–13	85	5.8	57
10–11	80	2.8	24
8–9	76	1.6	21
6–7	72	0.5	7
4–5	68	0.2	3

SOURCE: W. J. McNamara. Retraining of industrial personnel. *Personnel Psychology*, 1963, *16*, 233–247.

whether the characteristics be height, weight, or scores on a test, assumes a shape which approximates quite closely a bell-shaped curve or distribution. This is known as the *normal curve*. Because of this isomorphic relationship between real-world data and a mathematical equation that describes a theoretical curve, the theoretical curve can be used as a model for these data. Thus one might speak of test performance for factory workers as being, in general, *normally distributed*, implying that the general shape of the distribution of scores approximates the normal curve. Many statistical techniques so important to the industrial psychologist depend upon the principle of normally distributed scores.

Of course, not all frequency distributions are as symmetric as in Figure 2.2. Sometimes, conditions affecting job performance can cause distributions to change shape radically. For example, consider the curves shown in Figure 2.3. The first distribution represents the "true" frequency distribution one might find on a job if everyone performed at or near his normal capacity. The other two distributions illustrate how the shape can be influenced by such factors as work restriction or unlimited incentive. The middle curve has a shape that is *negatively skewed* while the bottom curve is *positively skewed*. Skewed distributions of test results do occur. For example, very difficult tests usually result in positive skewness and very easy tests result in negative skewness.

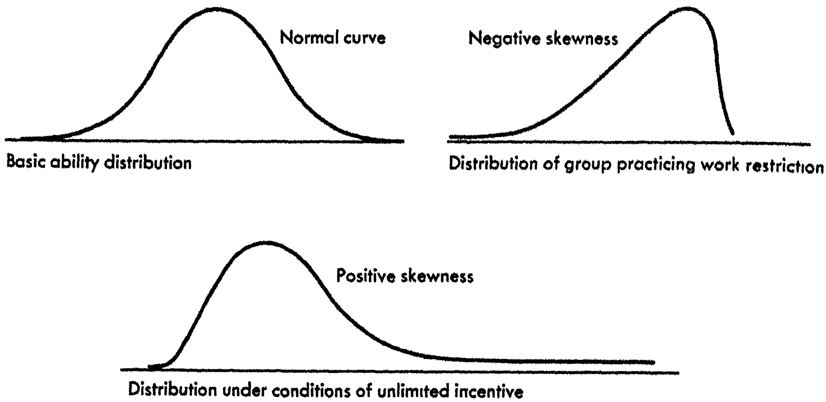


Figure 2.3. Distribution of scores of actual performance on a job may differ as a function of specific job conditions independent of the basic ability distribution.

BASIC SELECTION MODEL

The theory of testing in industry is based upon the individual differences among workers. Indeed, the purpose of the selection process is to take full advantage of such differences in order to select primarily those persons who possess the greatest amount of attributes important for success on the job.

Figure 2.4 diagrams the basic selection model. Each of the steps shown is important in developing a proper selection procedure.

Step 1: Examination of the job(s) having vacancies. This step consists of job analysis and will be discussed in greater detail in Chapter 17. Job analysis is a most critical and basic problem area in industrial psychology and is, or should be, the foundation of any industrial testing. A thorough knowledge and understanding of a job is of paramount importance and must precede the use of any test in the selection and placement of workers.

Step 2: Selection of criterion and predictor. The second step involves two parts—choosing an indicator which measures the extent of how "good" or successful a worker is (typically referred to as the *criterion*) and choosing a particular measure that can be used to predict how successful a worker will be on the job (typically referred to as the *selection device* or *predictor*). Criterion selection is an involved but basic problem; the solution decides how job success is defined and/or measured. Chapters 6 and 7 discuss some of these problems.

On the predictor side, the psychologist finds a wide variety of potentially useful devices which can be successful in discriminating between "good" and "poor" workers. Often used are such predictors as tests, interviews, application blanks, and letters of recommendation, among others. Chapters 4 and 5 present a more thorough look at various kinds of predictors and their relative merits.

Step 3: Measurement of performance. Once the criterion and the predictor have been selected it is necessary to obtain measures on both from a sample of workers on the job. This can be done either by giving the predictor to present

employees and simultaneously obtaining criterion measures, or by giving the predictor to new hires and waiting a specified time before obtaining the criterion measures (thus allowing enough time for the new hires to establish themselves as successful or unsuccessful). Both methods are used, and each has its advantages and disadvantages. These will be covered in more detail when we discuss kinds of validity.

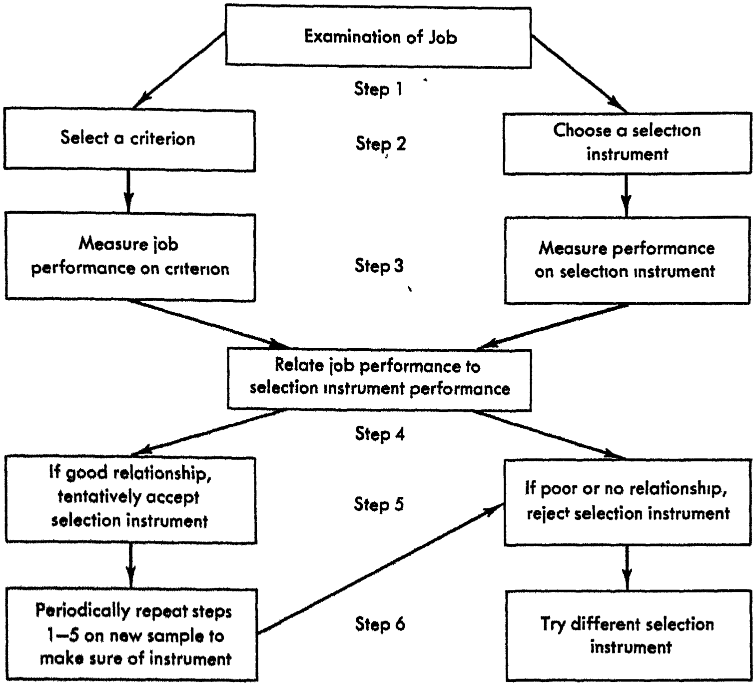


Figure 2.4. Basic selection model.

Step 4: Relating predictor to criterion. The fourth step in the selection process involves determining whether a true and meaningful relationship exists between the employee scores on the predictor and the criterion. Only if such a relationship exists can the selection process be considered successful. Establishing the existence of such a relationship is called *assessing the validity* of a predictor. This is usually a statistical process which involves the use and understanding of correlational methods and significance tests.

Step 5: Deciding upon the utility of the selection device. Making the final decision as to whether to use the predictor to select new job hires depends not only upon the size of the relationship found (in step 4) and its significance, but also upon many other conditions: the number of applicants, number of job openings, proportion of present employees considered successful (the *base rate*), and respective variances of the successful and unsuccessful worker groups. All of these additional aspects concerning the use of a predictor are discussed in later portions of this chapter.

Step 6. Reevaluation. The fact that the predictive situation is a dynamic, ever-changing one should never be forgotten. What makes for good selection today may not be at all appropriate tomorrow; applicants change, jobs change, and employment conditions change. Thus any good selection program should be reevaluated periodically to make certain it is doing the job for which it has been designed.

CORRELATION

Prediction of job success involves determining the extent to which the predictor is related to the criterion. For example, suppose one were interested in setting up a selection program to hire new file clerks. Suppose further that it had been decided to use a paper-and-pencil test of clerical aptitude as a potential predictor of file clerk efficiency, and that efficiency was to be determined by ratings of supervisors. Table 2.3 shows some hypothetical data for this assumed situation, giving scores for twelve file clerks on both the clerical test and the efficiency criterion measure. Figure 2.5 shows a graph of the data in Table 2.3.

TABLE 2.3 *Scores on Clerical Test and Job Proficiency Criterion for 12 File Clerks*

Clerk	*Clerical Test*	*Job Proficiency*
1	10	3
2	60	6
3	20	2
4	50	7
5	80	9
6	10	2
7	20	4
8	50	5
9	60	10
10	40	5
11	70	8
12	30	4

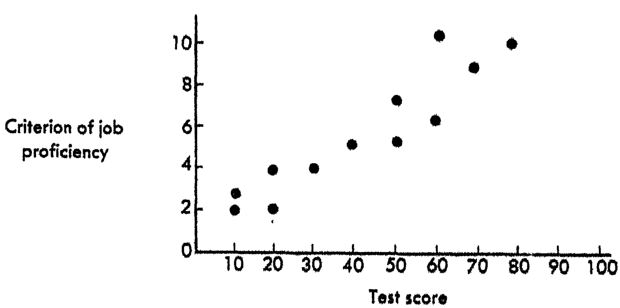


Figure 2.5. Plot of data in Table 2.3.

Notice that there appears to be a systematic trend. In general, the higher a person scored on the clerical test, the higher she scored on the measure of job proficiency. We can therefore deduce that there is a definite *relationship* between test performance (the predictor) and job proficiency (the criterion). We can also deduce that if we select those people who score higher on the test, we are more apt to hire people who will be more proficient than if we hire people independently of test score.

ESTABLISHING THE DEGREE OF RELATIONSHIP

The degree of relationship between any two variables may be defined as the extent to which these two variables vary together in a systematic fashion. The more technical term for this is the degree of *covariance* existing between variables. A formal measure of the degree of covariance between any two sets of scores is provided by a statistic known as the coefficient of correlation. When two sets of scores are highly related, we say they are highly *correlated*. The most common measure of correlation is the Pearson Product Moment Correlation Coefficient which is designated by the symbol r.

As a measure of relationship, r varies between $+1.00$ and -1.00. When r is $+1.00$, the two sets of scores are *positively* and perfectly related to each other. When r is -1.00, the two sets of scores are *negatively* and perfectly related to each other. When $r = 0.00$, the two sets of scores have no relationship to each other at all. Figure 2.6 shows graphs of different magnitudes of r.

In predicting job success the *sign* of the correlation coefficient is not important, but the magnitude is. The larger the absolute size of r, the better the prediction of criterion scores on the basis of information obtained from the predictor.

To understand the rationale of correlation it may be helpful to consider a pictorial representation of covariance and its relationship to r. Any set of scores will possess some amount of variation—in fact, as we have already seen, the scores of people on many traits follow a normal distribution with a small number of very

high scores, a small number of very low scores, and most of the scores occurring in the middle of the distribution. Suppose we represent this variance in a set of criterion scores as shown above, where the total area is defined as being 1.00. We can do this since it is possible to transform any set of raw scores so that their variance becomes equal to 1.00 using what is known as a z score transformation.

Similarly, suppose we have a set of predictor scores which also vary and are normally distributed, and again the area is defined as being equal to the quantity 1.00.

We can now represent r geometrically as being related to the amount of overlap (covariance) of the two sets of scores. A more precise definition of r as a statistic

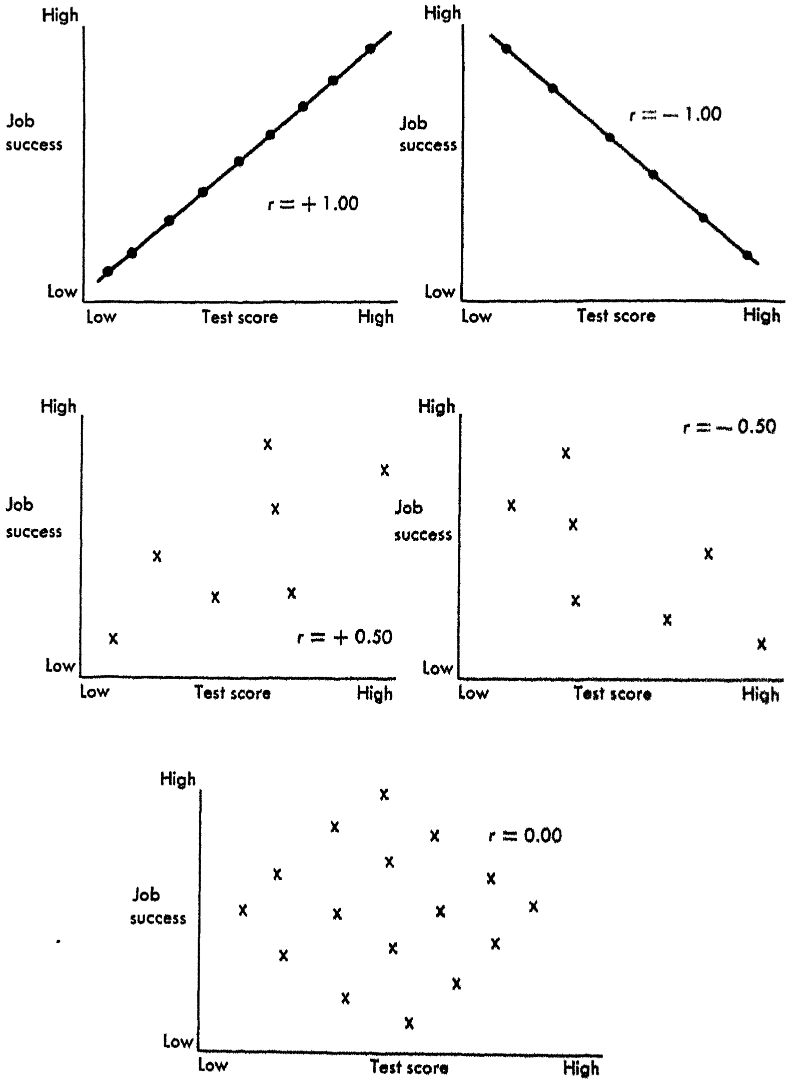
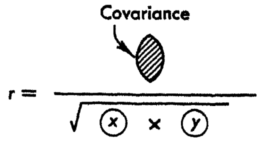
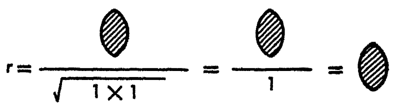


Figure 2.6. Correlation plots for various degrees of relationship between test score and job success.

is that it is the ratio of the amount of covariance between two variables to the square root of the product of the respective variances (sometimes called a geometric mean) which can be diagramed as shown below:



But when both x and y are equal to 1.00, we have r being equal simply to the amount of overlap between the two variables, as



Returning to the data given in Table 2.3, it is possible to compute the correlation between these two sets of scores using the formula

$$r = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2][N\Sigma y^2 - (\Sigma y)^2]}}$$

The reader is advised that r cannot be interpreted as a percentage. If $r = 0.50$, this does not imply that 50 percent of the variance in the criterion is predictable from the selection variable. The square of r, however, can be so interpreted. A correlation of 0.50, when squared, gives $r^2 = 0.25$, which may be interpreted as the percent of variance in the criterion predicted by the selection variable.

The statistic r^2 is sometimes called the *coefficient of determination* because it

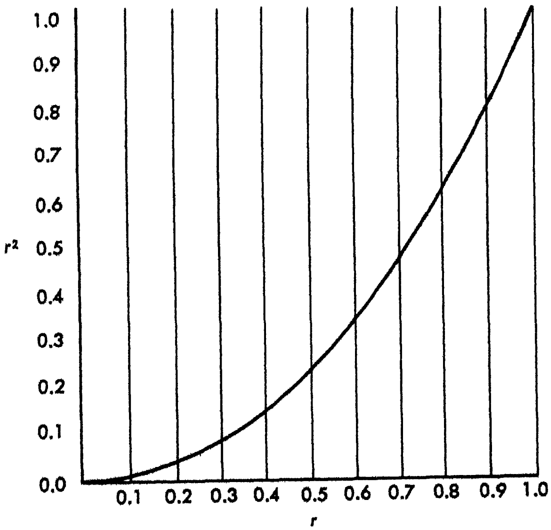


Figure 2.7. Relationship between r and r^2.

represents the amount of variance in one variable which can be "determined" by knowing the scores on a second variable. Figure 2.7 shows the relationship between r (the measure of relationship) and r^2. Note that it is possible to obtain r's of rather substantial size and still account for only a small proportion of the criterion variance.

REGRESSION

As we have seen, the correlation coefficient r measures the degree of relationship between two variables. By itself, however, it does not provide us with a procedure by which we can *predict* one set of scores from another set. The technique by which this is done is called *regression analysis*. Regression may be thought of as being related to correlation as follows: *Correlation* measures the magnitude or degree of relationship between two variables, while *regression* gives a description of the type of relationship between variables which in turn can be used to make predictions.

To illustrate regression, consider the scores plotted in Figure 2.8a. Obviously there is a substantial positive relationship existing between the predictor and the criterion in this case. Unfortunately, Figure 2.8a does not provide us with any information about the exact relationship other than the fact that it is a linear one (r always measures only the degree of linear, as opposed to curvilinear relationship, between two variables). If we want to predict criterion scores from some selection device, it is clear that we need to describe the observed relationship between predictor and criterion more specifically. This is accomplished by finding the line or function that best describes the data points. This is called fitting a "line of best fit" to the data. Since we are assuming the relationship to be linear (we used r to measure its magnitude), the type of line we use must be straight, that is, no curved lines are permitted. This best-fitting straight line is called the *regression line* and can be used to predict the criterion from the predictor.

Figure 2.8b shows two different lines of best fit which might be obtained if we asked two different persons to examine the data and then draw a line through the points which in their opinion seems best to describe the trend or relationship between the variables. While the general trend is similar, we find that the two people do not completely agree in their estimate of the relationship. This disagreement would in turn result in disagreement in the *predicted* criterion score depending upon which estimated regression line was used. Given a job applicant with a score x on the selection instrument, we would predict a criterion score of y_1 for this applicant if we were to use the first person's regression line; if we used the second person's regression line we would predict y_2 as the most likely criterion score. Which regression is correct?

This is a difficult question to answer unless there is some basis for deciding what a "best fit" really is. Fortunately, statisticians have generally agreed that a best-fitting line is one which goes through the points so that it minimizes the sum of the squared distances (in the y dimension) of the points from the line as shown in Figure 2.9.

A line which accomplishes minimizing Σd_i^2 is called a "least-squares" regression line. Such regression lines are mathematically directly related to r. Using the least-

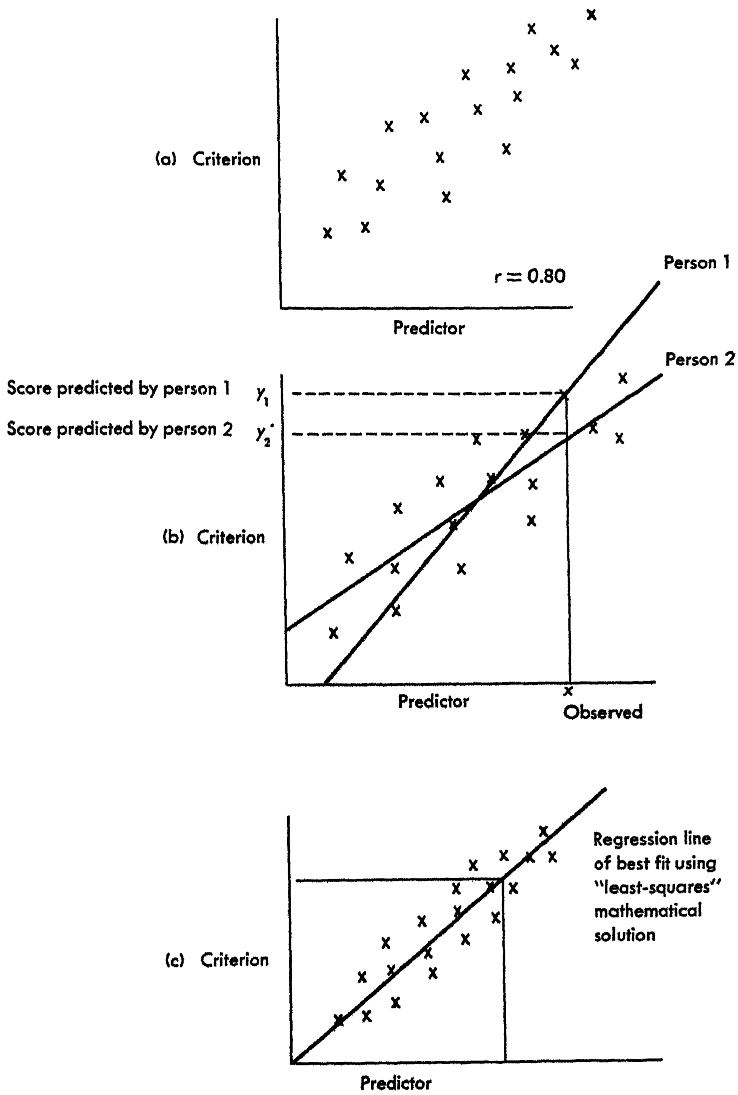


Figure 2.8. Finding a "line of best fit" to a set of points.

squares method for obtaining our prediction line will assure that different people will end up with the same line (assuming they make no errors in calculation). Similarly, the predicted criterion score for any particular x value will not vary depending upon who fits the prediction line (see Figure 2.8c).

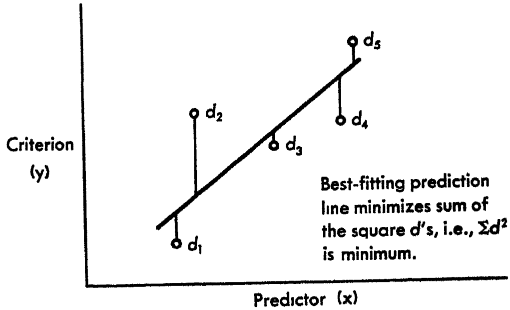


Figure 2.9. Example of "errors" in prediction for a line of best fit.

At this point the reader may ask, "Why do we need to predict criterion scores when we already have them?" The answer is quite simple. The initial measurement of the extent of the relationship between the predictor and criterion obviously requires both sets of scores or else the relationship could not have been established. Should the selection device prove useful, it can then be used with all new applicants for whom there can be a predictor score but for whom a criterion score does not exist. Our objective is to predict the criterion performance of future applicants. If a new applicant scores high on a test that was found to have a high positive relationship to the criterion, then we should expect him to have a high probability of turning out to be a successful hire.

STATISTICAL SIGNIFICANCE

Selection research always requires a knowledge and application of the basic concept of *statistical significance.* This basic concept allows us to state, with a specified likelihood of being correct, whether our findings represent the "true" state of affairs or whether they are simply an artifact of the general laws of chance. Expressed another way, tests of significance provide the researcher with a means of determining if the correlation coefficient is simply a chance phenomenon or if it adequately reflects the "true" correlation coefficient.

If a test of statistical significance indicates the obtained correlation is, in all likelihood, not due to chance sampling factors, then one may proceed to use the selection device with a certain degree of confidence that it will indeed perform successfully the task of discriminating good from poor performers on the criterion measure. On the other hand, if the test of significance indicates that the relationship between predictor and criterion is probably due to chance factors, the psychologist certainly has no justification for using the selection device to decide the fate of future job applicants—a fact that is sometimes tragically ignored by either

naive or unscrupulous individuals working in the area of selection and placement in industry.

It is important to understand that tests of significance are essentially indicators of chance probability. For example, *how sure must one be that results are not due to chance?* Generally speaking, most researchers accept with confidence results that could only be due to chance 5 times or less out of 100. They assume the risk of being wrong in their decision 1 time out of 20 by this process. This is known as adopting a "5 percent level of significance" in one's research. More cautious researchers will adopt even more stringent significant levels such as the 1 percent level. This means that they will only accept as nonchance those outcomes which have a chance likelihood of 1 time or less out of 100 cases. It is standard practice for all researchers to specify the level of significance which they have adopted for their research, as we will see in later chapters.

CHARACTERISTICS OF PREDICTORS

The two paramount requirements for any predictor are *validity* and *reliability*. In the industrial setting different kinds or types of validity exist, although the one that is most preferred is called *predictive validity*.

There are also different types of reliability measures. The concern with reliability and validity is not limited to predictors but applies to criteria as well—however, the discussion of the criterion is deferred to Chapter 6.

VALIDITY

The validity of a predictor can generally be defined as *the extent to which the predictor achieves certain aims of the user* by measuring what is supposed to be measured. The specific kind of validity involved thus depends upon the particular aim of the user in any situation.

PREDICTIVE VALIDITY The aim of the user is to employ his measuring instrument to predict the future performance of employees on some other variable (criterion). Predictive validity is established statistically through correlation and regression. The important distinction of predictive validity is one of a time element. Predictor scores are obtained on individuals at one point in time (e.g., time of hire) and criterion scores are obtained at a later date (e.g., at the end of six months). The resulting relationship thus truly represents the "predictive" power of the instrument. Predictive validity is the most important type of validity in selection since it is the only type which truly duplicates the selection situation. Another name which is sometimes used for predictive validity is *follow-up validity*.

CONCURRENT VALIDITY The aim here, at least in theory, should be to estimate the *present* performance of employees on some criterion measure from scores on the predictor. Concurrent validity is also established using correlation and regression techniques, but with no time lag between the obtaining of predictor and criterion scores. A sample of current employees is used to ascertain the predictor-criterion relationship, and then the resulting regression may be applied by obtaining predictor scores on the remaining job holders. In other words, we are interested

in predicting the *present* status of people, not their status at some future time. It is extremely important to point out that *high concurrent validity does not assure high predictive validity.* Unfortunately, concurrent validity is too often used in industry as a substitute for predictive validity Management is sometimes unwilling to wait for the time required by the predictive method, and may not realize that present employees can represent a basically different population of workers from job applicants. Workers who are presently employed have survived screening in both hiring and continuity, and the poorer workers who were hired may have left either voluntarily or by request. This makes it very difficult to justify generalizing concurrent validities to a predictive validity situation.

CONTENT VALIDITY When the validator assumes that his predictor is representative of a given class of situations, he is involved in content validity. He has a specific notion about the kind of knowledge, skill, attitude, or performance that should be tapped by the measuring instrument, and he considers the instrument valid to the degree that its content is representative of what he wants to tap. Content validity is generally not measurable in any statistical or quantitative sense. One finds the greatest use of content validity among users of achievement tests, such as final exams in a college course. A final exam could only be considered to have content validity if it adequately represented (sampled), in terms of its items, the content of the course. If it did not represent a coverage of course material, it certainly could not be considered an appropriate test to use for a final exam—i.e., it would not have content validity.

CONSTRUCT VALIDITY With this type of validity the user wishes to infer the degree to which the persons being evaluated possess some trait or quality (construct) presumed to be reflected in test performance. The general procedure involves administering several test instruments which logically appear to measure the same construct, and then observing the relationships between these measures. Construct validity has not been used to any considerable degree by the industrial psychologist, it tends to be more often used in theoretical rather than pragmatic situations.

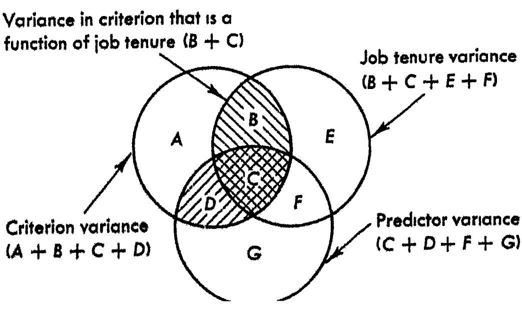
SYNTHETIC VALIDITY One might consider synthetic validity to be "assumed" predictive validity. Suppose we have a test that in a number of situations has demonstrated a high predictive relationship to various performance criteria of industrial foremen. Suppose further that a small manufacturing plant wants to use a test in selecting foremen, but too few foremen work in the plant to carry out even a concurrent validity study. This plant might decide to use the test without any formal statistical evaluation on the assumption that it had been demonstrated as being successful in other, larger plants. This procedure could only be considered valid if (1) the foreman's job in this plant is similar to the foremen jobs involved in the statistical evaluation of the test, and (2) the foreman applicants at this plant are typical (come from the same population) as the applicants for the foremen jobs in the larger plants. Synthetic validity should only be substituted for predictive validity with the full awareness of its possible limitations.

FACE VALIDITY Another kind of validity often used to describe a test involves the degree to which a user is interested in having his test "look right" to the test taker. Job applicants often become upset if the prediction instruments they are

required to take appear to have little or no relationship to the job for which they are applying. For example, if one is selecting people for a machinist position and a test of arithmetic ability is used as a predictor, the test items should deal with numbers applied to mechanical problems rather than being worded in more general terms such as the buying of apples or oranges. If the applicant fails to see the relevance of the predictor to the job for which he is applying, as often happens on personality tests, he may undergo a serious loss in motivation in the test situation, become derisive, or, on the other hand, feel insecure. This not only damages the selection program but also may harm the image of the company and damage the image of tests in any industrial setting. The authors would hazard a guess that some of the bad publicity received by users of selection devices in industry can be due to the user overlooking the need for his tests to have face validity.

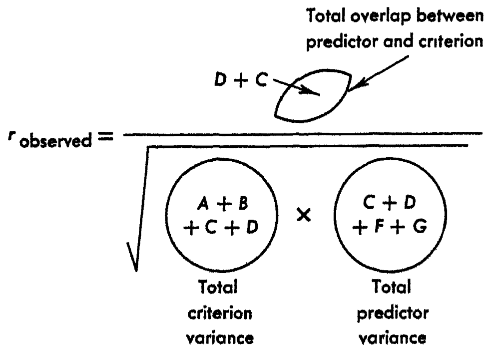
AGE, TENURE, AND JOB EXPERIENCE AND THEIR EFFECT ON VALIDITY Examination of the performance of workers on any particular job often shows a definite relationship between such variables as age and experience and the criterion. The more complex the job, the more these kinds of relationships are likely to exist. For many jobs a substantial amount of experience is necessary before employees become proficient at their work. The correlation between these kinds of variables and criteria of job success presents a serious problem in selection. Caution is necessary, particularly if one uses the concurrent validity procedure as a means of establishing the utility of any prediction device.

If, for example, there exists a high correlation between the criterion and length of time on the job, how should a high concurrent validity coefficient be interpreted? Does this mean that the predictor is truly reflecting ability differences among workers as measured by the criterion, or are worker differences primarily due to experience on the job? If it is the latter, then all the predictor is going to accomplish is to differentiate those workers with long tenure from those who have been hired more recently. The observed validity is generally an overestimate of the predictive efficiency of the selection instrument. In fact, unless one can clearly demonstrate that the predictor is *not* correlated with traits such as age and tenure which may themselves be determiners of job performance, all concurrent validities obtained with that predictor must be highly suspect. To illustrate the point, consider the situation where one has a criterion, a predictor, and a criterion-related variable such as job tenure which is largely responsible for the differences in skill shown on the criterion by employees, as follows:

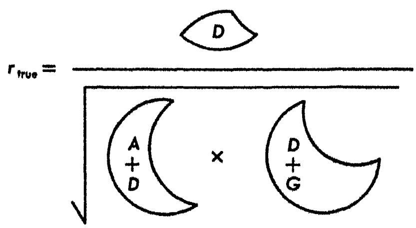


$C + D$ = Observed concurrent validity of predictor
D = Amount of "tenure-free" criterion variance accounted for by predictor
C = Amount of "tenure-determined" criterion variance accounted for by predictor

The observed validity is generally but not always an overestimate of the true validity, since:



The true or unbiased concurrent validity, which represents the correlation between predictor and criterion that is completely free of the influence of job tenure, is given by the equation—



The correlation (r_{true}) as shown in the diagram actually represents, in pictorial fashion, what is known in statistics as a "partial" correlation coefficient. It reports the correlation between predictor and criterion after the effects of job tenure have been removed from both the predictor scores and the criterion scores of present employees. It is important that tenure effects be removed from both criterion and predictor in the concurrent situation. If these effects are not statistically removed from the *criterion*, we will end up predicting the influence of tenure rather than job performance, with little or no relevance to predictive validity. If job tenure effects are not removed from the *predictor*, we may also obtain a validity coefficient which cannot be considered relevant to any truly predictive validity situation.

Certainly the problems of criterion and predictor correlated variables in the concurrent setting illustrate some of the serious constraints involved in this method of validation. It can safely be stated that there is absolutely no equal substitute for the type of validity known as predictive validity when constructing and using a selection instrument.

RELIABILITY

In general terms, the validity concept deals with what is being measured by a measuring device. A second and perhaps equally important characteristic of predictors is the need to know the consistency of the measure, regardless of what is being measured. Stated another way, we need to establish the degree of stability of any measuring device; the measurement obtained from a predictor must be consistent. The degree to which any measuring instrument is consistent or stable and would yield the same scores again and again if necessary is defined as the *reliability* of that test instrument.

Like validity, reliability is usually measured by means of the correlation coefficient. Since reliable measurement implies stability from one situation to another, reliable instrument should produce either the same scores or at least similar rankings of individuals in two situations. By computing the correlation we obtain a mathematical expression of the extent to which that occurs. Thus a reliable measuring instrument is one on which individuals receive the same score (or nearly the same) in repeated measurements. When the correlation coefficient is used to measure the similarity of scores for a group of people on two applications of the same measure, it is called a *reliability* coefficient.

The actual process by which one can assess the reliability of a measure depends upon numerous factors. There are three major alternative "kinds" of reliability, each of which has its separate advantages and disadvantages. They are sufficiently different in their underlying logic to justify examining each in some detail. The three techniques for obtaining instrument reliability are. (1) repeated measures on the same people with the same test or instrument, (2) measurement on the same people with two "equivalent" forms of the measuring instrument, and (3) separation of the measuring device into two or more equivalent parts and intercorrelating these "part" scores.

Before considering each method, we should examine in a more specific manner certain kinds of reliability or stability of measurement that we might conceivably be interested in under different circumstances. Let us assume that any time we use a measuring instrument to obtain a person's score, the score received is a function of several factors, as follows:

$$x_i = x_{\text{true}} + x_{\text{error}}$$

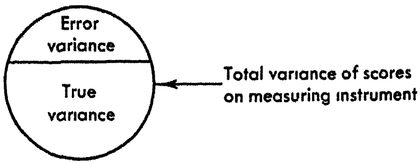
where x_i = observed score for person i on test

x_{true} = true score for person i on test—this is the actual amount of quality measured by the test that person i really possesses

x_{error} = error score for person i on test—this is the amount that person i's score was affected by the operating of various chance or time factors.

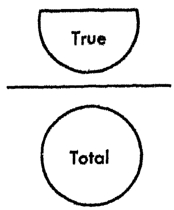
If all measuring instruments and measuring methods were "error-free," then we would always obtain the true scores of people, and the correlation between two measurements on the same group of people would always be + 1.00 or perfect reliability (assuming no change in the true scores is to be expected). Unfortunately, such error-free measurement is never completely available, since a wide variety of things contribute to performance at any particular moment in time. Thus, x_i may

either be greater than or less than x_{true} for any particular measurement, and correlations computed between measurements are always less than unity. In terms of our pictorial representation of the variance of performance among people on any measuring device, whether it be test or interview, predictor or criterion, this total variance may be divided into the two major components of true variance and error variance:

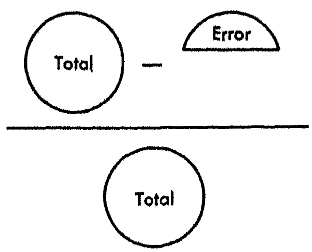


where total variance = total variability of observed test scores
true variance = variability of people in terms of their true amounts of the characteristic being measured
error variance = variability of people's error scores

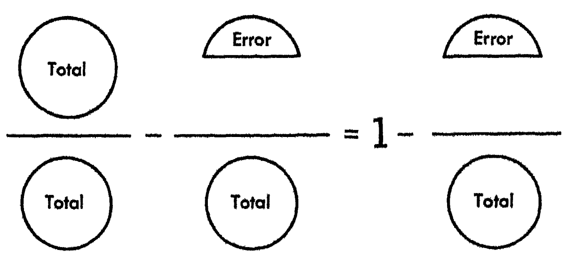
Reliability may be defined as a ratio of true variance to total variance, or



thus,



which can be expressed as



The larger the proportion of true score variance, or conversely, the smaller the amount of error variance present in the measuring process, the greater the reliability of measurement.

The critical factor differentiating the three major procedures for determining reliability is in the process of deciding what is to be considered error variance and what is to be considered true or systematic variance. There is no single reliability for any test. Rather, the reliability will depend upon the needs of the moment. For example, the psychologist might ask any of the following kinds of questions about the measurement process:

1. How accurately can I measure people with this test at any given moment in time?
2. How accurately will measures taken with this test today be representative of these same people at some future point in time?
3. How accurately will the scores on this test represent the true ability of these people on the trait being sampled by the test?

All three are legitimate reliability questions. However, each places a somewhat different emphasis on various sources of error variation in test scores. These sources of error variation have been expressed by Thorndike and Hagen (1963) as:

1. Variation due to the test at a particular moment in time
2. Variation in the individual from time period to time period
3. Variation due to the particular sample of tasks chosen to represent the quality being measured

Let us now proceed to examine each reliability method, keeping in mind the error sources so we can determine how each method treats each source.

TEST-RETEST METHOD One obvious method to evaluate stability consists of measuring the same individual's performance twice with the same measuring instrument. This type of reliability includes variation sources 1 and 2 as error. Thus the resulting reliability is one which measures the stability of the true score over time. There are numerous problems with the test-retest method that are created by having the individuals measured on the *same* test twice. For example, unless the time period is fairly long between administrations, the variable of a memory factor is likely to bias the responses of people on the second administration. Another difficulty is that variation due to the particular sample of tasks or items chosen is treated as systematic variance which adds to the reliability.

Thus any person who by chance happened to know more answers simply because a few of the test items touched, say, upon a hobby of that person, would also be favored in the second administration because the same items, rather than a new sample, is used. He should therefore score high on both testings due to variation source 3 being treated as true variance.

PARALLEL TESTS METHOD One way of avoiding having error source 3 as true variance is to use two completely comparable or "equivalent" forms of the measuring instrument. These two forms should be as identical as possible, except that *specific* items or questions on each form would *not* be the same although they would each represent a *similar sample of items chosen.* One form may be admin-

istered immediately after the other or they may be administered at spaced intervals, depending upon whether one is concerned with having variation source 2 included as error variance.

This kind of reliability, when spaced testing is used, represents the most rigorous evaluation of stability that can be made. However, it is often impossible or at best extremely difficult to construct alternate forms of a measuring instrument.

How does one construct two alternate but equivalent forms of a measure of job performance or two alternate forms of a personal history form? In many cases, not without considerable difficulty. This lack of a truly comparable measuring device has caused psychologists to look for additional methods of assessing reliability besides the test-retest and parallel form procedures.

SUBDIVIDED TEST METHOD The third major reliability method is often referred to as a measure of the *internal consistency* of a measuring device. It provides an indication of the extent to which people score the same, relative to each other, on different subdivisions of the overall instrument. This method is probably the most widely used method of measuring reliability since it requires that only one form be constructed and yet does not require repeated administrations of that form.

Its mechanics are very simple. In its most basic form, the internal consistency method is the parallel forms procedure in which the parallel forms are two halves of the same test. These half-tests are selected so as to be as equivalent as possible, although often the test is simply divided into two halves by putting all the odd-numbered items into one half and all the even-numbered items into the other half. This is called the *odd-even* version of the *split-half* technique.

It is important to remember that the separation of the total test into equivalent halves occurs only when scoring that test—not when administering it. Since the two subtests are each only half as long as the original, each represents a sample of behavior only half as large as the total test. Thus the correlation (reliability) between halves is likely to be an *underestimate* of the reliability of scores based upon the entire test. To obtain an estimate of what the reliability of the complete test is, the Spearman-Brown Prophecy formula may be applied as follows:

$$r_{tt} = \frac{2r_{\frac{1}{2}\frac{1}{2}}}{1 + r_{\frac{1}{2}\frac{1}{2}}}$$

where r_{tt} = reliability of the total test (estimated)

$r_{\frac{1}{2}\frac{1}{2}}$ = observed correlation between the two halves of the test.

For example, if the observed correlation between halves were 0.40, then the Prophecy Formula would estimate the reliability of the complete test as:

$$r_{tt} = \frac{2\,(0.40)}{1 + 0.40} = \frac{0.80}{1.40} = 0.57$$

The split-half method thus provides a method for estimating reliability with a single test and a single administration. There are, however, certain drawbacks to its use. When one has a test that involves primarily speed factors (such as certain simple clerical tests), the alternate-half procedure gives a spuriously high result.

Since speed tests usually involve easy items, it is only a question of whether they were responded to that determines whether they were correct or incorrect. Thus, splitting the test on an odd-even basis, for instance, would result in virtually identical scores for both halves—thus a high positive correlation.

KUDER-RICHARDSON METHOD Another version of the split-half method is frequently used for measuring reliability. Related to a statistical technique known as the analysis of variance, its most frequent form is known as the Kuder-Richardson procedure. The Kuder-Richardson (K-R) method is also an internal consistency reliability which essentially treats *each test item* as a subtest, so that instead of having two halves there are n subtests, where n is the total number of items on the measuring instrument. The K-R technique is equivalent to computing all possible correlations between pairs of tests items (there will be $n[n-1]/2$ such pairs), taking the average of these, and adjusting the result by using the Spearman-Brown prophecy formula:

$$r_{tt} = \frac{K\bar{r}_{it}}{1+(K-1)\bar{r}_{it}}$$

where r_{tt} = estimated reliability of total test
$\bar{r}_{it}$ = average correlation between items
K = number of pairs of items

Like the split-half forms procedure, the Kuder-Richardson procedure ignores variation source 2 and is not appropriate for speed tests.

A summary comparison is given in Table 2.4. This table shows the various reliability methods and compares them in terms of the kinds of variation they include as error variance.

TABLE 2.4 *Sources of Variation Represented in Different Procedures for Estimating Reliability*

	Experimental Procedure for Estimating Reliability					
Source of Variation	*Immediate Retest (Same Test)*	*Retest After Interval (Same Test)*	*Parallel Test Form Without Time Interval*	*Parallel Test Form With Time Interval*	*Odd-Even Halves of Single Test*	*Kuder-Richardson Analysis of Single Test*
How much the score can be expected to fluctuate owing to:						
Variations arising within the measurement procedure itself	X	X	X	X	X	X
Changes in the individual from day to day		X		X		
Changes in the specific sample of tasks			X	X	X	X

SOURCE: Adapted from R. L. Thorndike and Elizabeth Hagen. *Measurement and evaluation in psychology and education*, 2nd ed. Wiley, New York, 1963, 182.

DETERMINING THE UTILITY OF A SELECTION INSTRUMENT

The *utility* of a prediction device is the degree to which its use improves the quality of the people being selected beyond what would have occurred had that device not been used. So far we have discussed validity and reliability, both of which play an important role in determining the utility of any selection instrument. In addition there are, however, several other factors which are equally critical in determining utility in any situation involving group selection. These additional variables are (1) criterion reliability, (2) criterion relevance, (3) the selection ratio, and (4) the percent of present employees considered successful.

The reader is cautioned to keep in mind that group prediction is the process of selecting systematically a subsample of applicants who are more apt to succeed, on the average, than either the group as a whole or any subgroup selected at random from the whole. This differs from the process of individual prediction where one is concerned with predicting the likelihood of success of a *particular person,* rather than a group of persons. There are, of course, selection situations that involve both group and individual aspects of selection. One example is a selection program used by the United States Navy in their selection of flight cadets. The Navy uses group prediction techniques prior to and during the several different phases of flight training. In addition, it is also necessary for predictions to be made for a specific cadet and his individual likelihood of success in the program. The need for these latter kinds of predictions occurs whenever a cadet's record during training has been poor enough to bring him before a review board. For this Navy program the same basic predictors are used for both the group and the individual predictions.

The reliability of the criterion, a topic that has been deferred to the criterion chapter (Chapter 6), does not differ in method or logic from predictor reliability; it is usually assessed by one of the previously discussed reliability techniques. Criterion relevance, another topic discussed in greater detail in Chapter 6, is the criterion equivalent of validity but is usually *not* a formal statistical process.

PREDICTOR VALIDITY

The primary statistical index influencing the utility of any prediction instrument is its validity. While it will be demonstrated that low and even zero validity situations can still result in quite successful selection under special conditions, the validity coefficient remains the central variable in selection. To illustrate, consider the dia-

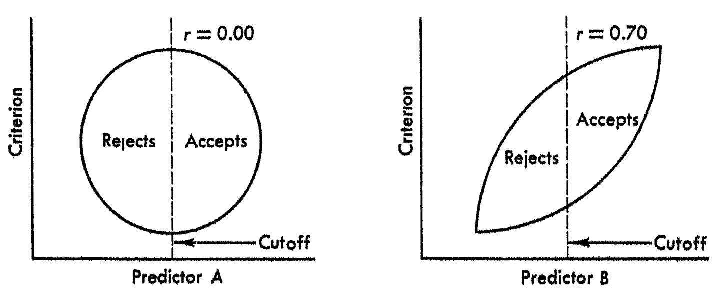


Figure 2.10. Two different predictor-criterion relationships.

grams shown in Figure 2.10 in which two different predictor-criterion relationships are shown, one with a validity of 0.00 and the other with a validity of 0.70. In both cases a cutoff score on the predictor has been established that allows us to take the top 50 percent of the people taking the test. Which predictor will result in the greatest increase in average criterion score of the selected group over what has been achieved by prior methods (that is, random selection)? Looking first at predictor A in terms of how the people are distributed on the *criterion dimension only*, we find that the mean criterion score of the "accepted" group is exactly the same as the "rejected" group. That is, the people who are accepted by taking the top half of the scores on test A do *not* tend to have higher criterion scores than do the lowest 50 percent of the scorers on test A, as shown in Figure 2.11.

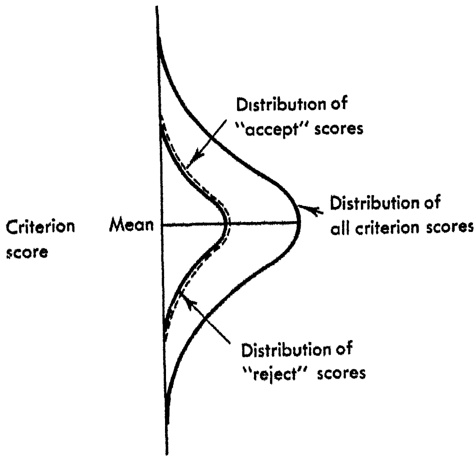


Figure 2.11. Frequency distribution of criterion scores for (1) total group, (2) accepted group, and (3) rejected group using predictor A

However, when we look at predictor *B* we get quite a different picture. We can see immediately that those people above the cutoff on *A* seem to do better on the criterion than do those below the cutoff. That is, the people above the cutoff have a higher *average criterion score* than do those below. This is shown in Figure 2.12, which again shows the three distributions of criterion values.

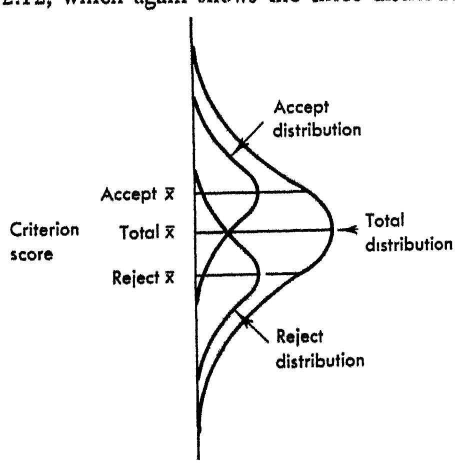


Figure 2.12. Frequency distribution of criterion scores for (1) total group, (2) accepted group, and (3) rejected group using predictor *B*.

Thus we appear to have our first general principle in test utility: given any arbitrarily defined cutoff on a test, the higher the validity, the greater the increase in average criterion score for the selected group over that observed for the total group. In other words, the difference

$$(\bar{x}_{\text{selected group}}) - (\bar{x}_{\text{total group}})$$

will increase in direct proportion to the test validity. Indeed, it can be shown algebraically that this is so (later we will see certain exceptions to this first principle).

Recently Naylor and Shine (1965) have published a set of tables which offer easy computation of the increase in average criterion score which will be achieved with any test given that the test validity and test cutoff point can be specified. This table is given in the Appendix along with explanations and examples of its use.

SELECTION RATIO AND PERCENT OF SUCCESSFUL EMPLOYEES

Two other variables which play an important role in determining the utility of a predictor are the selection ratio and the percentage of present employees considered successful. The reader will recall that the utility of a predictor was defined as the improvement in quality of the hires obtained using a prediction device when compared with present methods of selection. Quality is typically defined in terms of (1) the average criterion score of the group, or (2) in terms of the proportion of people in that group who have criterion scores above some value that is considered to be minimal in order for one to be a successful employee. For any given obtained validity coefficient between criterion and the predictor, a manipulation of either the selection ratio and/or a change in the percent of present employees considered successful will result in marked changes in the resulting quality of the hired (selected) employees.

SELECTION RATIO Simply described, the selection ratio (SR) may be expressed as:

$$n/N = \text{SR}$$

where n = number of job openings
N = number of job applicants available for placement

When the SR is equal to or greater than 1.00, the use of any selection device has little meaning. With more job openings than job applicants, the applicant is in a seller's market where the company may need to purchase his services regardless of his quality. If, however, the SR is less than 1.00, then there are more job applicants than positions and the employer is in a position to be selective in terms of whom he hires.

The way in which the SR can influence the selection process can best be demonstrated by referring to Figure 2.13. In Figure 2.13a, a scatterplot of scores is shown that is approximately the shape that could be expected with a large sample of people and a correlation between predictor and criterion of 0.70 (the higher the correlation, the more nearly the scatterplot will approach a straight line; the lower the correlation, the more nearly the scatterplot will approach a circle). The proportion of the oval that is shaded represents the proportion of applicants who are actually

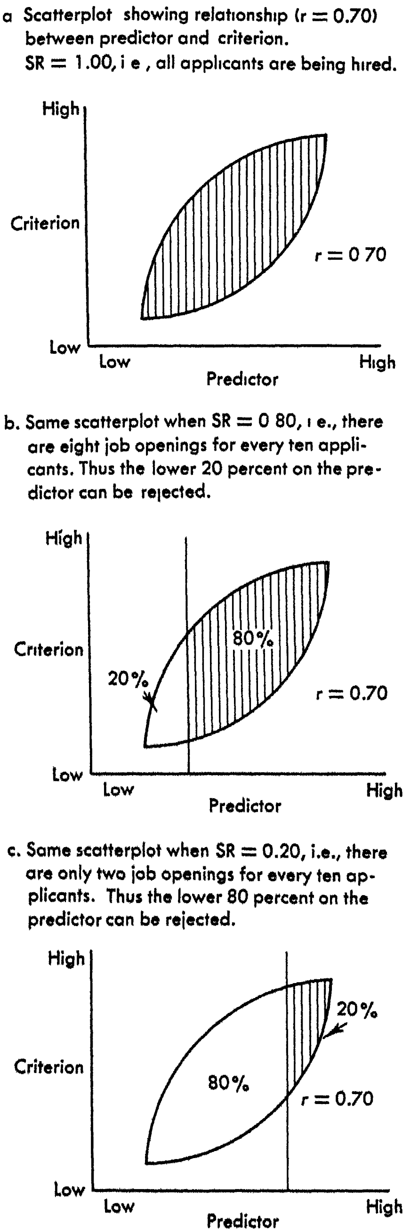


Figure 2 13. Diagrams showing the effect of selection ratio on the average quality of those being hired.

hired, *viz.*, the SR. In Figure 2.13a, an SR of 100 is presented; there is a job opening for every applicant so all will be hired.

In part b of Figure 2.13 we see what happens to the average quality of those being hired when the SR becomes 0.80. Since there are jobs for only 80 percent of the applicants, the employer will logically hire the 80 percent having the highest predictor scores, since the predictor is highly related to subsequent criterion performance. These 80 percent are represented by the shaded area of the oval falling to the right of the cutoff point on the predictor. Since those being eliminated generally possess low criterion scores, it is easy to see how the average criterion score for those hired with an SR of 0.80 is higher than it is if a random group of applicants were placed on jobs as in Figure 2.13a. This increase in average quality is shown even more dramatically in Figure 2.13c which illustrates an SR of 0.20. Faced with a situation where there are ten applicants for every two jobs, the employer is "sitting pretty"—he now can select the top 20 percent of the performers. These individuals are represented by the shaded area of the oval falling to the right of the cutoff in Figure 2.13c. The difference in average quality of criterion score for this select subgroup as opposed to that of the entire group is very large. The benefits to the employer in terms of dollars in this situation should certainly be substantial.

The general principle that a lower selection ratio will always result in better quality employees being hired holds as long as the relationship between the predictor and the criterion is some value greater than zero (negative or positive *r*'s are equally effective if of equal magnitude). In fact, it can be demonstrated that the principle of the selection ratio can be effectively utilized in some cases even if *all* applicants need to be hired. This can occur if there are at least *two* jobs, each with a number of openings and each of which has its own predictor with greater than zero validity.

PERCENT OF PRESENT EMPLOYEES WHO ARE SUCCESSFUL In our discussion concerning validity and the SR we have so far assumed the criterion to be continuous and therefore the higher the criterion score, the more *satisfactory* that worker is considered to be. Let us now suppose a criterion score exists that defines whether a worker is either satisfactory or unsatisfactory—that is, if he performs above a standard he is considered satisfactory and if he performs below this standard he is considered unsatisfactory. The diagrams in Figure 2.14 illustrate this. In part a, a relationship of about 0.70 between the criterion and the predictor is shown. Note that the horizontal line, called the *criterion cutoff*, separates all workers into two groups: those considered successful and those considered unsuccessful. Such a cutoff will, of course, have to be rather arbitrary in its nature. However, in many cases it is not too difficult to arrive at some consensus concerning minimal acceptable performance.

Part b of Figure 2.14 shows the same data with a predictor cutoff based upon a selection ratio of about 0.5. The last part of the figure shows both cutoffs together. When combined in this fashion it becomes possible to distinguish among the various subportions of the data which are formed by the intersection of the two cutoff lines.

Part A. Those applicants who are to the right of the test score cutoff and above the criterion cutoff are called *true positives*. They are those who the test says ought to be successful and who indeed will be successful according to the criterion. They represent correct decisions based upon the test.

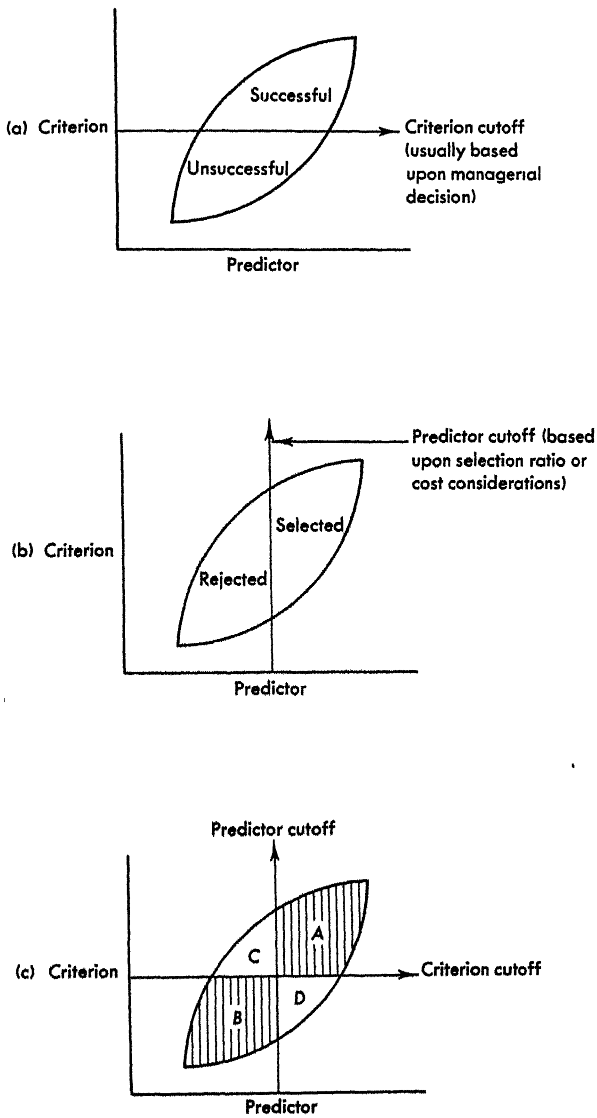


Figure 2.14. Effects of establishing criterion and predictor cutoff points on a bivariate distribution of scores.

Part B. This segment includes those applicants who have scores below the predictor cutoff and below the criterion cutoff. Called the *true negatives*, these applicants, like the true positives, represent correct decisions based upon the predictor.

Part C. These applicants have scores *below* the predictor cutoff but *above* the criterion cutoff. These people would not be hired if hiring decisions were based upon the test, in spite of the fact that their eventual criterion score was high enough to place them in the satisfactory category. This represents one kind of mistake or error that occurs in testing and is referred to as *false negatives*.

Part D. The last segment of the oval consists of job applicants who would be hired but who would subsequently turn out to be unsatisfactory in their work. These persons also represent "mistakes" in the selection process and are known as *false positives*.

Several meaningful ratios may be constructed using the various parts of Figure 2.14c. For example,

$$(1) \qquad \frac{C+D}{A+B}$$

This is a ratio of the number of errors in selection to the number of employees correctly placed. The size of this ratio depends upon all three variables: the location of the criterion cutoff, the location of predictor cutoff, and the validity coefficient. Not only is the size of this ratio affected by these variables, but so is the relative magnitude of the two types of errors, *C* and *D*. Usually the employer is more concerned with minimizing fake positives than he is concerned about the number of false negatives. This often is seized upon by those opposed to testing as being one of the major evils of scientific selection via tests, namely that some people are rejected who would be successful on the job if given an opportunity to prove themselves. The reader will have to deliberate the pros and cons of this problem for himself—the authors merely point out the difficulty. However, the authors hasten to add that industrial psychologists can be as socially minded as their critics. Industrial psychologists generally have the data to tell the whole story, while some critics without any data at all merely "holler" about one error.

Another ratio of importance is given by

$$(2) \qquad \frac{A+C}{A+B+C+D} = \text{percent presently successful}$$

This represents the percentage of the present employees who are satisfactory. It is a base percentage that expresses the degree of success that is obtained with whatever selection methods were used prior to the introduction of the predictor. The third ratio,

$$(3) \qquad \frac{A}{A+D} = \text{percent successful using predictor}$$

is an expression of the proportion of *hired* applicants who will be successful if one uses the predictor as an aid to selection along with the methods currently being employed. To the extent that (3) is greater than (2), the predictor is adding something to the selection process. In comparing the relative magnitude of (2) and (3), some general principles may be stated:

1. For any particular validity and criterion cutoff, a reduction in the SR will cause an increase in the *effective* validity. Thus, one can compensate for low statistical validity if one can be selective in his hires.

2. For any particular statistical validity and selection ratio, the smaller the percentage of present employees considered satisfactory, the larger the percentage *increase* of satisfactory applicants obtained using the predictor. In other words, if we define the difference between ratios (2) and (3) as

$$\text{Utility} = \frac{A}{A+D} - \frac{A+B}{A+B+C+D} = \text{percent increase in effectiveness}$$

where effectiveness is defined as the percent of successes being employed, then the greatest benefit will be observed under those conditions where the poorest job is currently being done—a logical outcome. There are some exceptions, of course. For example, consider Figure 2.15.

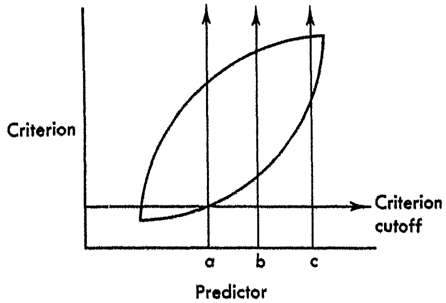


Figure 2.15. Effect of different selection ratios in a situation where the majority of present employees are considered satisfactory

Notice from Figure 2.15 that no matter which of the three different selection ratios one uses, 100 percent of all hired applicants will eventually be judged satisfactory. Thus here is one situation where large selection ratio changes are of no consequence.

268236

TAYLOR-RUSSELL TABLES

A detailed expression of the exact relationships between the size of the validity coefficient, the selection ratio, and the percent of presently satisfactory employees has been prepared by Taylor and Russell (1939). Under given conditions of validity, selection ratio, and percent satisfactory, their tables allow one to determine the percentage of hires who will be satisfactory using the predictor in conjunction with current methods. However, the Naylor-Shine tables discussed in the section on predictor validity would appear to have several advantages over the Taylor-Russell tables. The Naylor-Shine tables are formulated in terms of differences in average criterion score between the selected group and the original group; Taylor and Russell use differences in the percent successful between the selected group and the original group. Thus, the Naylor-Shine tables would seem to give a more meaningful index of test utility. Also, the use of the Taylor-Russell tables requires that the employees be separated into two groups, "successful" and "unsuccessful," by selecting some arbitrary point on the criterion dimension that represents "minimal satisfactory performance." The Naylor-Shine tables do not require any decision of this sort for their use and are therefore more general in their applicability.

A Note of Caution. Both the Naylor-Shine tables and the Taylor-Russell tables have certain limitations which are *very* important. Both methods for evaluating test utility are based upon the assumptions that (1) the relationship between predictor and criterion is a linear one, and (2) the validity coefficient used is one obtained by concurrent validity procedures.

Smith (1948) and others have pointed out the dangers that exist if one tries to use tables such as Taylor and Russell's under conditions where the relationship is not linear between the predictor and the criterion. Such a relationship is shown in Figure 2.16. When such nonlinear relationships exist, both tables are completely inappropriate for determining test utility.

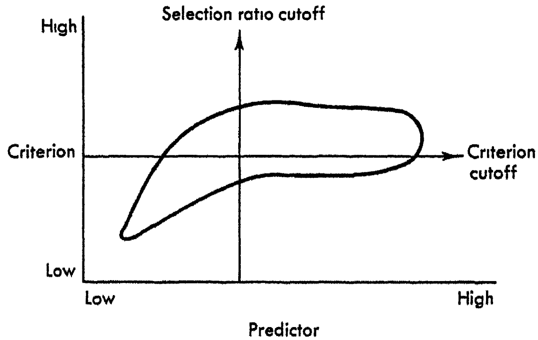


Figure 2.16. Example of nonlinear scatterplot—a situation which makes the Naylor-Shine and Taylor-Russell tables inappropriate.

The fact that both tables assume a validity coefficient based upon concurrent validation procedures may come as a surprise since earlier it was pointed out that concurrent validity was not a particularly good substitute for predictive validity. However, test utility involves determining the increase in either average criterion score (Naylor-Shine tables) or percent of successful employees (Taylor-Russell tables) over that *currently being obtained with present employees.* The basic scatterplot is one based upon present employees hired by the normal selection procedures —the typical concurrent validity paradigm.

RELIABILITY OF PREDICTOR AND CRITERION

Reliability of the criterion and of the predictor are also important, primarily because they influence or put limits upon the size of the validity coefficient that may be obtained. There is a basic algebraic relationship which exists between validity and the reliability of the predictor and the criterion which is

$$r_{pc\,(\text{obtained})} = r_{pc\,(\text{true})} \sqrt{r_{pp} \times r_{cc}}$$

where $r_{pc(\text{obtained})}$ = observed correlation (validity) between the predictor and criterion

$r_{pc\,(\text{true})}$ = "true" correlation (validity) between the predictor and criterion

r_{pp} = reliability of predictor

r_{cc} = reliability of criterion

Notice from the above relationship that only when r_{pp} and r_{cc} are unity (perfect reliability) will the obtained validity be equal to the true validity. As the reliability of the two measures decreases, so will the obtained validity. For example, suppose $r_{pc_{(\text{true})}} = 0.60$, $r_{pp} = r_{cc} = 0.80$, then $r_{pc_{(\text{obtained})}} = 0.60 \sqrt{0.80 \times 0.80} = 0.60 (0.80) = 0.48$. Also notice that if the reliability of either the predictor or criterion is zero, then the obtained validity will be zero as well.

CRITERION RELEVANCE

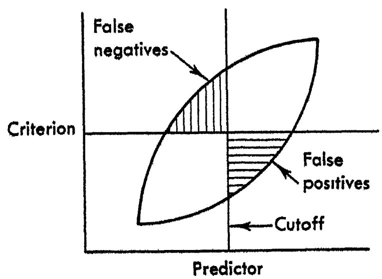
The relevance of a criterion has little to do with the actual empirical utility of a prediction instrument, although it has a great deal to do with its logical utility. This topic will be discussed further in Chapter 6.

A DECISION-THEORY APPROACH

The problem of selection can be viewed from a somewhat different perspective than the one used in the earlier portions of this chapter. This second approach proves interesting in that we shall find that predictor *validity* may not be as important a variable in selection as the traditional viewpoint makes it out to be. Our new perspective is one based upon a decision theory model. We should begin by restating the objective in a typical selection situation. *In many selection situations we wish to establish a cutting score on our predictor which will result in minimizing our decision errors.*

Implicit in this type of situation is the assumption that the selection ratio can be manipulated at will; that is, it is not "fixed" at some value. Also implicit is the notion that our criterion variable can be meaningfully separated into two or more distinct groupings such as "successful" and "unsuccessful." Our objective is to manipulate the cutting score (which is the same as manipulating the selection ratio) in order to *minimize the number of errors made in our process of deciding whether a person should be hired or rejected.*

Earlier we pointed out that there were two distinct types of decision errors in the selection paradigm, false positives and false negatives, as shown below:

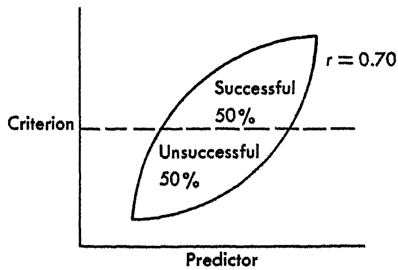


Our objective then is to find the cutoff point which will result in the smallest number of total errors. For purposes of convenience we shall start by assuming that both types of error are considered *equally costly*. That is, we have no reason to prefer making a false positive error over a false negative error, or vice versa. By

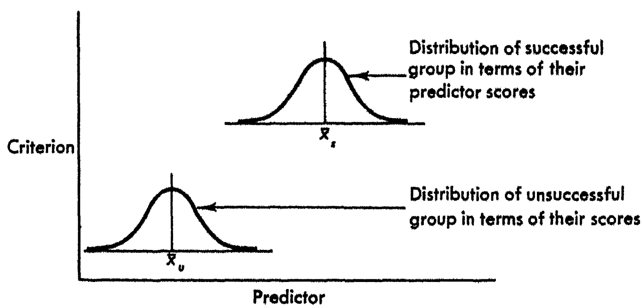
making this assumption it is possible to cast the problem directly in terms of minimizing the total number of both kinds of errors rather than having to weigh the two types of errors by their respective "cost."

LOCATION OF THE CUTOFF POINT

To illustrate how the problem of finding an optimum location for our cutting score can be approached, consider the case where we have a specified validity (e.g., about 0.70) and a specified percent of present employees considered successful (often referred to in this context as the "base rate"). This can be diagramed as follows:

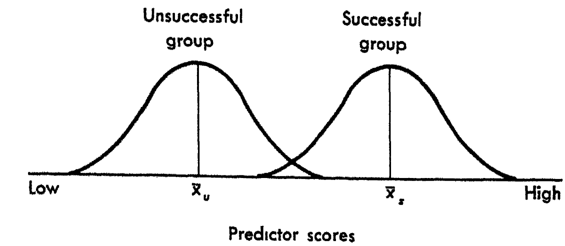


The next step is to present the same data in a slightly different form. First, we know that our total group of employees is assumed to have a normal distribution in terms of their predictor scores. Second, and equally important, *both* subgroups (successful and unsuccessful) are assumed to have normal distributions. By looking at the example above it is easy to deduce that the *mean predictor* score of the successful group is going to be higher than that of the unsuccessful group. We might diagram this as:



Both distributions will be equal in size since they are based upon the same number of people (i.e., 50 percent in each group). There is an algebraic relationship between the difference between the means of the two subgroups as viewed in this fashion and the size of the correlation coefficient. If the group means are significantly different from each other (say at a significance level of 0.05), then the correlation coefficient will also be found to be significant at the same level.

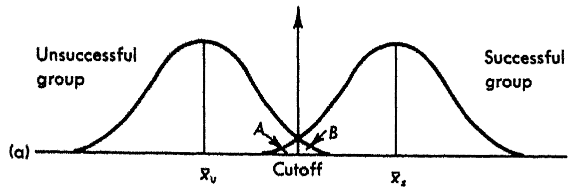
Taking our diagram one step further, we can place the two frequency distributions of the subgroups side by side on the same baseline, as shown below.

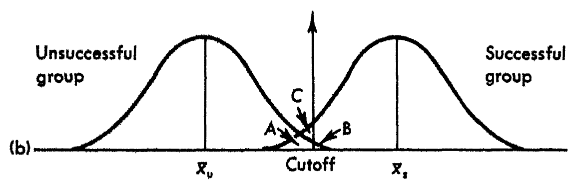


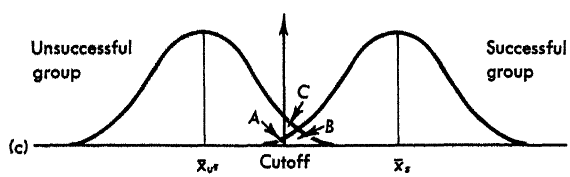
After doing this we can now return to our original question—where do we locate a cutoff on the predictor so that the total number of errors will be minimized?

It turns out that the mathematical solution to this problem results in a very simple answer: *The cutoff point which minimizes total error is the point at which the two distributions intersect each other.*

This can be easily demonstrated on a conceptual level by looking at the three cases illustrated below. The same difference between the means (that is, the same correlation) is used in each case—all that has been changed is the location of the cutoff point on the predictor.







In illustration (a), the number of false positives (failures who are above cutoff) is given by the area *B*. The number of false negatives (successes who are below the cutoff) is given by the area *A*. Thus,

$$\text{Total error} = A + B$$

For illustration (b), the number of false positives is given by B and the number of false negatives is given by $A + C$. Thus,

$$\text{Total error} = A + B + C$$

For illustration (c), the number of false positives is given by $B + C$ and the number of false negatives is given by A. Thus,

$$\text{Total error} = A + B + C$$

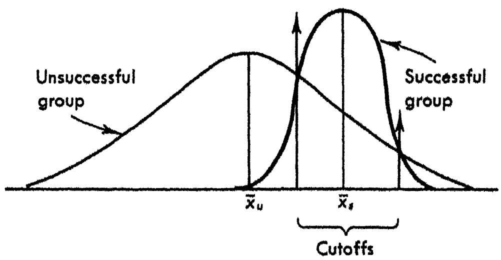
Since inspection of all three illustrations quickly confirms that the area $A + B$ is the same for all three cases, then it is obvious that error is increased by some amount C whenever the cutoff is moved *away* (in either direction) from the point at which the two distributions intersect each other.

SOME UNUSUAL RAMIFICATIONS

We now have a general principle for locating a cutting score that will minimize the total number of errors in a selection decision-making situation—namely, at the point of intersection. It turns out that, as long as both types of errors are considered equally costly, this is a very general rule and is *not* affected by (1) the relative sizes of the two groups (i.e., percent considered successful), or (2) the respective variances or dispersions of the two distributions.

This leads to some interesting and very important aspects of the general prediction problem concerning the relationship of test validity to test utility. Rorer, Hoffman, LaForge, and Hsieh (1966) have pointed out three such interesting cases.[1]

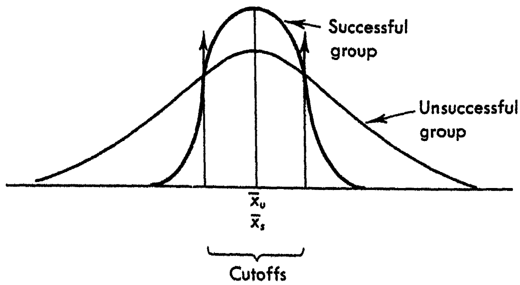
Case 1: Both the means and the variances of the two groups differ from each other. Suppose that our successful group is of equal size to the unsuccessful group and has a significantly higher mean on the predictor, but its variance is much smaller. A diagram of such a situation is as follows.



Our principle of establishing cutoff points says that we should place them wherever the two distributions intersect. Note that this happens *twice* in this particular case. Thus, we have an upper cutoff and a lower cutoff. We should select only those people who fall within the interval between cutoffs in terms of their test score. Any *other* cutoff points will result in greater total error than would be obtained with the ones located at the points of intersection.

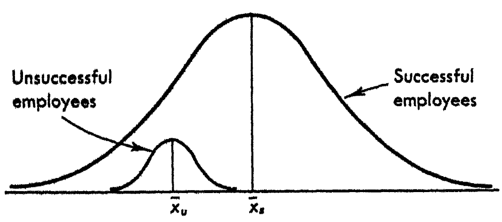
[1] It should be pointed out that the assumptions for the use of the correlation coefficient are violated in this case (unequal variances for subgroups), so that it should not even be used as a measure of relationship.

Case 2: Groups have equal means but different variances. In this very interesting case the two groups do not differ in terms of their mean predictor score—that is, *on the average* the unsuccessful employees do just as well on the test as do the successful employees. This implies that the correlation coefficient is zero between the predictor and the criterion. However, we have further stated that the two groups differ in terms of their variability. If we assume the successful group is the group with the smaller variability for purposes of exposition, we can express this diagrammatically as follows:



Even though the two groups have the same mean criterion score, it is possible to develop cutoff points which will improve prediction over that currently being enjoyed through present methods, since the two distributions intersect at two points due to their unequal variability. Thus, we have the unique situation where there would be no apparent validity (as measured by a correlation coefficient) but where prediction can be much improved by use of appropriate cutoffs [2]

Case 3: Group means are considerably different but group size is also greatly different. Suppose we are dealing with a situation in which the base rate of unsuccessful employees is very small, that is, about 90 percent of our present employees are considered successful. Such a situation is shown in the following diagram.



Here we have another unique situation. Even though the group means may be substantially different thus giving a substantial correlation between criterion and predictor, it is not going to be possible to establish *any* cutoff which will result in reducing error over what is currently obtained with present methods. Because of the marked difference in size between the two groups, we see that the two dis-

[2] It should be pointed out that again the assumptions for the use of the correlation coefficient are violated in this case (unequal variances for subgroups), so that it should not even be considered as a measure of relationship.

tributions do *not* intersect at any point. Under our present selection system we are only making errors 10 percent of the time. If we move our cutoff from left to right in case 3 (it is located in the extreme left to begin with, since we are currently selecting all these people) we will, of course, start eliminating some of the unsuccessful people currently being employed under the present system At the same time, however, we are going to start rejecting employees who would turn out to be successful. Looking at the diagram quickly tells us that this increase in false negatives would be greater than the corresponding decrease in false positives no matter where we put our cutoff. Thus, any test-based cutoff will result in more errors than we have without the test, even though the test is highly valid.

SOME CONCLUSIONS

The basic decision-theory approach to selection and placement outlined above has a number of advantages over the more classical approach based upon the correlation model. Recent work by Rorer, Hoffman, LaForge, and Hsieh (1966), Cronbach and Gleser (1965), Darlington and Stauffer (1966), and Mahoney and England (1965) indicates the growing interest in the decision-theory approach as applied to selection problems. There is no question but that it is a more general and better model for handling this kind of decision task, and we predict that in the future problems of selection and placement will be treated in this context more frequently—perhaps to the eventual exclusion of the more stereotyped correlational model.

REFERENCES

Cronbach, L. J., and G. C. Gleser. *Psychological tests and personnel decisions*, 2nd ed. University of Illinois Press, Urbana, Ill., 1965.

Darlington, R B , and G. F. Stauffer. Use and evaluation of discrete test information in decision making. *Journal of Applied Psychology*, 1966, 50, 125–129.

Mahoney, T. A., and G. W England. Efficiency and accuracy of employer decision rules. *Personnel Psychology*, 1965, 18, 361–377.

Naylor, J. C., and L. C. Shine. A table for determining the increase in mean criterion score obtained by using a selection device. *Journal of Industrial Psychology*, 1965, 3, 33–42.

Rorer, L. G., P. J. Hoffman, G. E. LaForge, and K. Hsieh. Optimal cutting scores to discriminate groups of unequal size and variance. *Journal of Applied Psychology*, 1966, 50, 153–164.

Smith, M. Cautions concerning the use of the Taylor-Russell tables in employee selection. *Journal of Applied Psychology*, 1948, 32, 595–600.

Taylor, H. C., and J. T. Russell. The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. *Journal of Applied Psychology*, 1939, 23, 565–578.

Thorndike, R. L., and E. Hagen. *Measurement and evaluation in psychology and education*, 2nd ed. Wiley, New York, 1963.

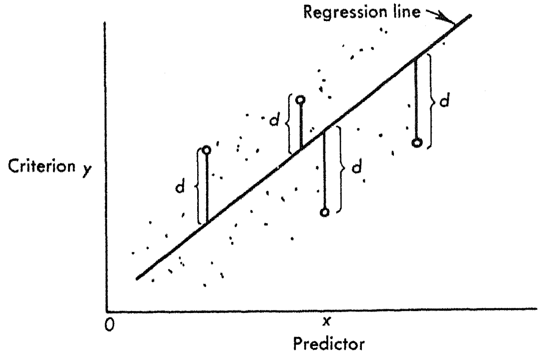


Figure 3.1. Best-fitting line for a set of points.

MULTIPLE PREDICTION

Before proceeding to an examination of the basic selection models which are available to the psychologist, it is necessary to concern ourselves with a brief look at the general multiple prediction model. This model is usually referred to as the *multiple regression model.* In Chapter 2 we assumed a linear world and said that in the general prediction paradigm we develop a regression line to fit the set of data points defined by people's scores on a predictor (the x-axis or abscissa) and on the criterion (the y-axis or ordinate). Figure 3.1 shows such a situation.

The regression line in Figure 3.1 is a straight line and is located so that the sum of the squared distances from each point to the line (running parallel to the y-axis) is as small as possible. We use a best-fitting straight line since we have assumed a linear relationship between x and y. The basic formula for a straight line is

$$\hat{y} = a + bx$$

where $\hat{y}$ = predicted score on criterion
a = a constant indicating the point at which the regression line crosses the y-axis
b = slope of the line, represented by $\Delta y/\Delta x$, or the change in y observed for a corresponding change in x
x = observed score on predictor

Thus, the basic regression line model appears as shown in Figure 3.2.

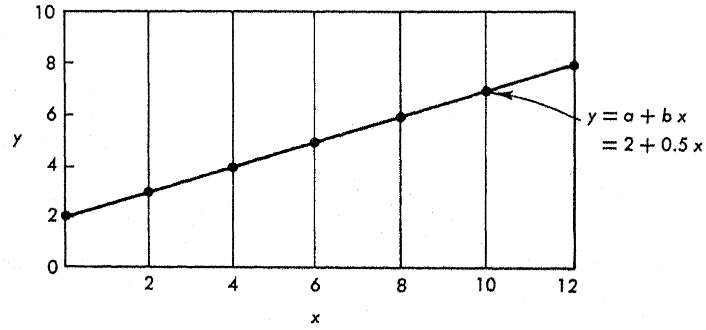


Figure 3.2. Example of a linear model.

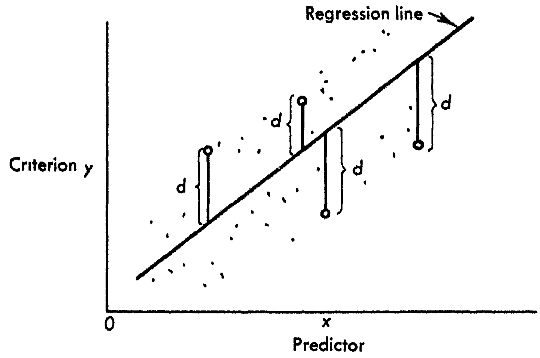


Figure 3.1. Best-fitting line for a set of points.

MULTIPLE PREDICTION

Before proceeding to an examination of the basic selection models which are available to the psychologist, it is necessary to concern ourselves with a brief look at the general multiple prediction model. This model is usually referred to as the *multiple regression model.* In Chapter 2 we assumed a linear world and said that in the general prediction paradigm we develop a regression line to fit the set of data points defined by people's scores on a predictor (the x-axis or abscissa) and on the criterion (the y-axis or ordinate). Figure 3.1 shows such a situation.

The regression line in Figure 3.1 is a straight line and is located so that the sum of the squared distances from each point to the line (running parallel to the y-axis) is as small as possible. We use a best-fitting straight line since we have assumed a linear relationship between x and y. The basic formula for a straight line is

$$\hat{y} = a + bx$$

where $\hat{y}$ = predicted score on criterion
a = a constant indicating the point at which the regression line crosses the y-axis
b = slope of the line, represented by $\Delta y/\Delta x$, or the change in y observed for a corresponding change in x
x = observed score on predictor

Thus, the basic regression line model appears as shown in Figure 3.2.

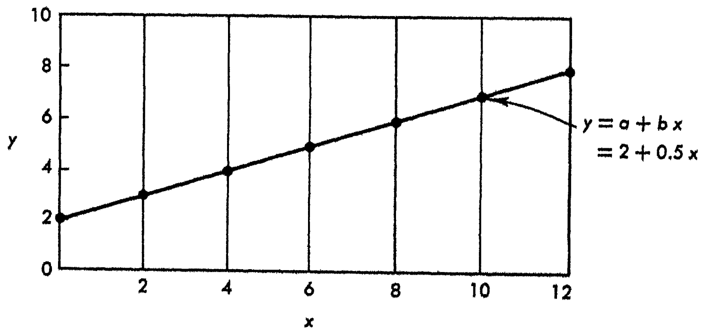


Figure 3.2. Example of a linear model.

Note that in Figure 3.2 the regression line crosses the y-axis at a value of 2 Thus $a = 2$. Also note that for every 2-unit increase in x there is a corresponding 1-unit increase in y. *Thus* $\Delta y/\Delta x = 1/2 = 0.5 = b$. The regression equation then becomes

$$\hat{y} = 2 + 0.5x$$

Given any x value, we have a regression line that allows us to predict a y score corresponding to it. For example, if x were 8, then

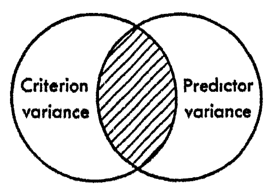
$$\begin{aligned}\hat{y} &= 2 + 0.5\,(8)\\ &= 2 + 4\\ &= 6\end{aligned}$$

To summarize: In the single predictor case, one computes a best-fitting straight line to the observed points, where the term "best fit" means the sum of squared deviations of the observed values around the line will be a minimum. The formulas necessary to compute the constants a and b which define this best-fitting line are called "least-squares" formulas and are as follows:

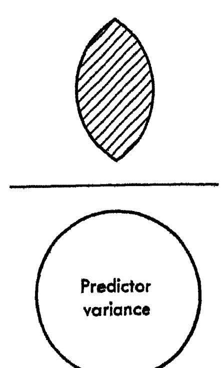
$$b = \frac{N\Sigma xy - (\Sigma x)\,(\Sigma y)}{N\Sigma x^2 - (\Sigma x)^2}$$

and $$a = \bar{y} - b\bar{x}$$

The formula for b is a ratio of the covariance between the predictor and criterion and the total variation in the predictor. Thus if we have



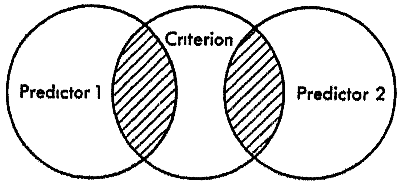
then b equals



and if the reader will refer back to the diagram on page 31 of Chapter 2 he will see that when the criterion variance and predictor variance are equal, then $b = r$, or the slope of the regression line equals the correlation coefficient.

TWO PREDICTORS

It is logical to assume that if predictor x_1 can contribute to the successful prediction of criterion scores, and if predictor x_2 can also contribute to the successful prediction of criterion scores, then using both predictors together should allow for better overall prediction than using either predictor individually. However, the degree to which the two predictors (when combined) will improve the predictability depends upon several factors, most important of which is the correlation between the two predictors themselves. Consider, for example, the situation where two predictors each correlate substantially with a criterion but do *not* correlate with each other, as follows:

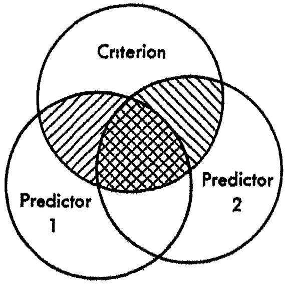


Clearly, a great deal of additional criterion variance can be explained using predictor 2 along with predictor 1. The combined relationship between two or more predictors and a criterion is called a *multiple* correlation and has the symbol R. As was the case with r^2, the value of R^2 represents the total amount of criterion variance which can be explained by using several predictors. When predictors 1 and 2 are not correlated with each other, the squared multiple correlation coefficient can be shown to be an additive function of the individual squared correlation coefficients, or

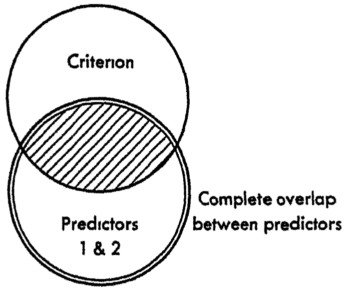
$$R^2_{c.12} = r^2_{1c} + r^2_{2c} \tag{3.1}$$

Thus, when r_{12} (the intercorrelation of predictors) is zero, then the squared multiple validity is the sum of the squared individual validities.

When two predictors *are* correlated with each other, things become somewhat more complex. Consider a situation (as in the following diagram) where each predictor has substantial individual validity but where r_{12} is also rather large.



Because of the intercorrelation between these predictors, the diagram shows that the amount of overlap between predictor 2 and the criterion can be divided into two parts: that area unique to predictor 2 and that area shared with predictor 1. Thus, the use of a second predictor in this situation allows us to account for more criterion variance than could be done using predictor 1 alone, but all of the criterion variance predicted by 2 is not new variance. A general rule can therefore be stated concerning multiple predictors. *All other things being equal, the higher the correlation between predictors, the less the overall prediction will be improved by using both predictors together.* The extreme case, of course, would be the situation where the predictors were perfectly correlated and we would have no additional criterion variance accounted for by the addition of predictor 2 to our selection battery.



In the case of two predictors which are correlated with each other, we can express R^2 as a function of the separate validities and the size of the intercorrelation between predictors with the formula[2]

$$R^2_{c.12} = \frac{r^2_{1c} + r^2_{2c} - 2r_{12}r_{1c}r_{2c}}{1 - r^2_{12}} \qquad (3.2)$$

note that if $r_{12} = 0$, then formula 3.2 reduces to

$$R^2_{c.12} = r^2_{1c} + r^2_{2c}$$

which is formula 3.1.

A more explicit illustration of the influence of predictor intercorrelation upon the size of the multiple correlation coefficient may be obtained from Table 3.1, where examples of R and R^2 values are given for pairs of predictors having validities of 0.30, 0.50, and 0.70 under hypothetical conditions of 0.00, 0.30, and 0.60 intercorrelation. Figure 3.3 shows the general trend using the data given in Table 3.1. The moral to the psychologist is quite evident—avoid using predictors which are apt to be highly related to one another.[3]

[2] $R^2_{c.12}$ should be read as "the squared multiple correlation" when predicting the criterion from predictors 1 and 2 together.

[3] A procedure which allows selection of the most efficient set of *k* predictors from a larger group of *M* predictors is available. The method is known as the Wherry-Doolittle Test Selectior Procedure.

TABLE 3.1 *The Effects of Correlation Between Predictors upon the Squared Multiple R Obtained Using Those Predictors*

r_{1c}	r_{2c}	r_{12}	r_{1c}^2	r_{2c}^2	R	R^2
0.30	0.30	0.00	0.09	0.09	0.42	0.18
0.50	0.50	0.00	0 25	0.25	0.71	0.50
0.70	0 70	0 00	0 49	0.49	0.99	0.98
0.30	0.30	0.30	0 09	0.09	0 37	0.14
0.50	0.50	0.30	0.25	0.25	0.62	0.38
0.70	0.70	0.30	0.49	0.49	0.87	0.76
0 30	0.30	0.60	0.09	0.09	0.34	0.12
0.50	0.50	0.60	0.25	0.25	0 56	0 31
0.70	0.70	0.60	0.49	0.49	0.78	0.61

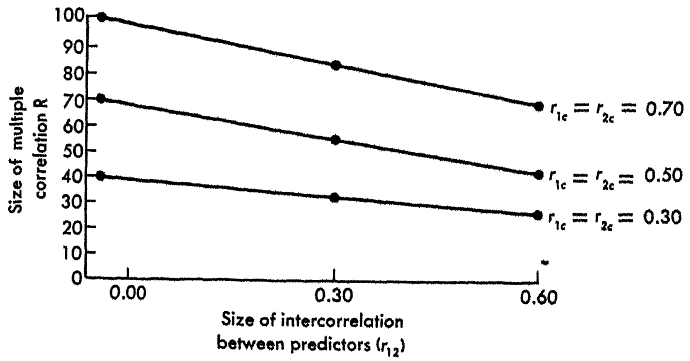


Figure 3.3. An illustration of the effect of predictor intercorrelation upon the multiple correlation coefficient.

PREDICTION EQUATIONS

The prediction equation in a two-predictor situation is an extension of the one-predictor model. The general form of the equation is

$$\hat{y} = a + b_1x_1 + b_2x_2 \tag{3.3}$$

This is the equation for a plane instead of a straight line. For the reader familiar with geometry, Figure 3.4 presents a three-dimensional drawing of the relationships between the variables x_1, x_2, and y corresponding to equation 3.3. Formulas are available which allow for computing the constants a, b, and b_2 which will result in the best-fitting regression plane. Once these constants have been determined, the resulting equation can then be used to make criterion performance predictions of new job applicants, given their scores on the separate predictors.

To illustrate, suppose data are available on 100 men hired for job X during a particular month which includes scores in two tests as well as criterion data after a six-month period. These data can be analyzed to determine the values for a, b_1,

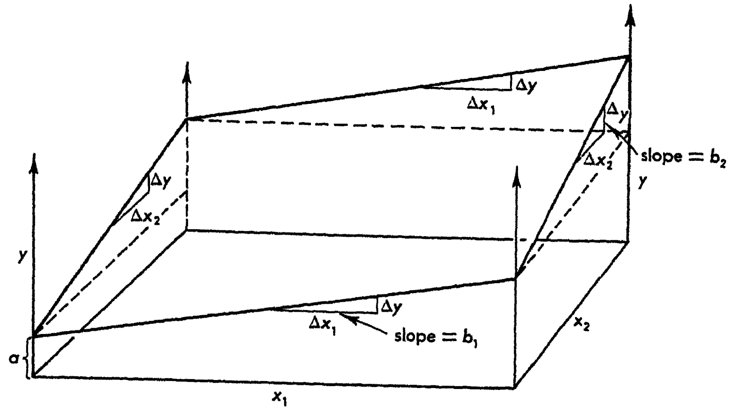


Figure 3.4. The geometry of a two-predictor regression equation. The tilted "top" of the box is the regression "plane."

and b_2 which best described the relationships between the variables. Suppose the following equation was the end result:

$$\hat{y} = 2 + 0.5x_1 + 0.9x_2 \tag{3.4}$$

This equation says that the most likely criterion score for any new hire will be equal to one-half his score on test 1 plus nine-tenths his score on test 2 plus two.[4] Thus if a new applicant scores 20 on test 1 and 30 on test 2, his predicted criterion performance at the end of six months from time of hire would be:

$$\begin{aligned}\hat{y} &= 2 + 0.5\,(20) + 0.9\,(30)\\ &= 2 + 10 + 27\\ &= 39\end{aligned}$$

The extension of the two-predictor model to a k-predictor model, where k is some large number of potential prediction of job success, is not too difficult conceptually. Our model expands to the form

$$\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3 + \ldots + b_kx_k \tag{3.5}$$

However, the computational procedures for solving for the least-squares values of all the constants in such an equation becomes rather complex unless one has computer facilities available. The reader is also cautioned to remember that in all the preceding discussion there has been the implicit assumption of a linear world, i.e., all the relationships between pairs of variables are linear ones. It is possible to modify the multiple regression model to avoid this assumption, but that is beyond the scope of this book.

PREDICTION SYSTEMS

By this time it should be clear that the typical prediction problem, be it selection, placement, or both, involves using a number of predictors. These predictors are used in the best manner possible as guides in making a decision about employment.

[4] We of course assume our new hire comes from the same population as the 100 men from whose data we derived our equation.

There are decisions such as "should he be hired for this job?" or "should she be sent to this training program?" There are several strategies that the psychologist can adopt in terms of his approach to the decision-making process. Depending upon the particular prediction system adopted, employment decisions may turn out quite differently. While each system has its own advantages and disadvantages, each provides a method for making decisions about people based upon a group of traits or qualities (the predictors) believed to be relevant to job success. The major systems are: (1) the *multiple regression system,* (2) the *multiple cutoff system,* (3) the *profile matching system,* and (4) the *multiple hurdle system.* Each system will be examined in the following sections in greater detail.

MULTIPLE REGRESSION SYSTEM

As the name implies, this placement system utilizes the multiple regression model for making decisions about individuals. The multiple regression model takes the form

$$\hat{y} = b_1x_1 + b_2x_2 \qquad \text{(assume } a = 0) \tag{3.6}$$

The use of such a model in selection assumes that (1) the traits x_1 and x_2 are linearly related to criterion performance, and that (2) possession of a "lot" of one of the traits compensates for having only a "little" of the second trait.

Given a situation, for example, where $b_1 = 2$ and $b_2 = 4$ and $a = 0$, the formula

$$\hat{y} = 2x_1 + 4x_2 \tag{3.7}$$

would be used for predicting job success. Let us suppose that a criterion score of 50 could be considered satisfactory performance by the employees and anything less resulted in nonsatisfactory performance. Table 3.2 shows some test scores on the two predictors for four theoretical job applicants. The predicted criterion score for each applicant has also been computed using equation 3.7. Note that all four applicants have exactly the same predicted criterion performance even though their test score patterns differ quite markedly. As we proceed from person A through person D, we see that their scores on test 2 systematically diminish. However, this drop is compensated for by a corresponding increase in test 1 performance. In fact, a close inspection will show that a gain of two points on test 1 is necessary to compensate for the loss of each point on test 2. This should not be surprising, since the relative weight given to test 2 is twice that given to test 1 in our regression model (i.e., $b_1 = 2$, $b_2 = 4$).

Figure 3.5 shows even more clearly the dynamics of the selection process created by the data in Table 3.2. The envelope of scores shown in the scatterplot of Figure

TABLE 3.2 *Predictor Scores and Predicted Criterion Scores Using* $\hat{y} = 2x_1 + 4x_2$ *for Four Theoretical Job Applicants*

Applicant	x_1 *Score on Test 1*	x_2 *Score on Test 2*	$\hat{y}$ *Predicted Criterion Score*
A	5	10	50
B	11	7	50
C	15	5	50
D	25	0	50

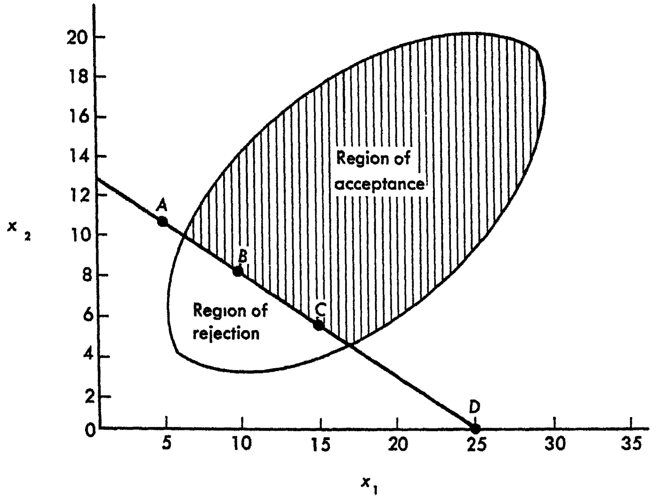


Figure 3.5. Regions of rejection and acceptance for job applicants formed by a situation where the regression equation is $y = 2x_1 + 4x_2$. Example shown illustrates a situation where x_1 and x_2 are moderately correlated with one another.

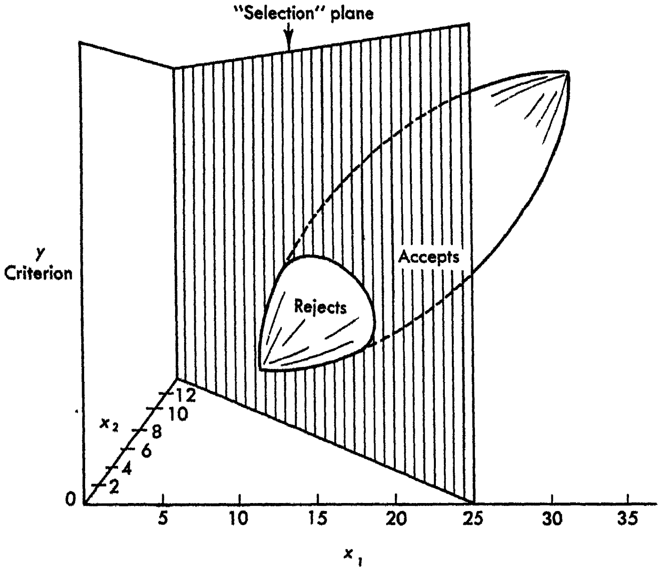


Figure 3.6. Extension of Figure 3.5 to three dimensions so that the observed criterion scores can also be shown. The scatterplot of scores, assuming linearity, will form a shape similar to that of a football. The rejection line is actually a plane that bisects this football into two parts. All people on the far side of the plane will have predicted criterion scores of 50 or better; all those on the near side will have predicted criterion scores of less than 50.

3.5 presents a situation where the two predictors of performance, x_1 and x_2, are positively correlated. If the correlation r_{12} were zero, the scatterplot would of course be a circle. However, the shape of the scatterplot is not critical to the trade-off concept inherent in the multiple regression system. Since we have said that any person with a predicted score of 50 or better was to be considered "satisfactory," we can plot the "50-point line" in Figure 3.5 which shows all possible combinations of test 1 and test 2 scores which will result in a criterion score of exactly 50 points using equation 3.7. As the figure indicates, all four applicants lie along this line.

An interesting aspect of Figure 3.5 is that the line divides the population of job applicants into two groups or regions. All applicants to the right and above the line are going to have criterion scores (using equation 3.7) which will be above 50. All applicants to the left and below the line will have criterion scores of less than 50. Thus, only the former will be accepted for employment since it is predicted that their performance will be satisfactory. The latter applicants, with predicted performance less than satisfactory, will be rejected with this selection system. Figure 3.6 extends Figure 3.5 into three dimensions, showing the observed criterion scores as well as the predictor scores for all individuals.

It is important to note that the plane in Figure 3.6 that divides the employees into those who would be selected using the multiple regression model given by equation 3.7 and those who would be rejected is *not* the regression plane. It is more properly called the *selection plane*. The reader is referred back to Figure 3.4 for an illustration of the regression plane in a two-predictor multiple regression system.

ASSUMPTIONS, ADVANTAGES, AND DISADVANTAGES OF MULTIPLE REGRESSION SYSTEM The multiple regression prediction system is a powerful selection procedure when used appropriately. Provided the basic assumption that all relationships are linear is true, it has a mathematical elegance that is hard to exceed. One knows, for example, that the model minimizes the errors in prediction. Another advantage of this system is that predictors are combined to obtain the most efficient estimate of subsequent performance.

One of the major points of controversy concerning the multiple regression model involves the trade-off principle so implicit in its use. Whether X units of one variable can be substituted for X units on another variable is always a moot question.

Certainly the method can be an extremely flexible one. It is possible to set up equations for each of a number of jobs using either the same or different predictors. As a result, predicted scores can be computed for each person for each job. People could then be hired and placed on a specific job using one or more of the following procedures:

1. Place each person on that job for which the predicted score is highest. This assumes the organization will profit most if each person is placed where he has the most aptitude, regardless of the absolute amount of that aptitude. If no positions are open in that job, he would be placed on another job for which he received the second best criterion score. One problem with such a procedure is that it is possible that the jobs themselves might have different minimum requirements for success. Thus, it could happen that his best score (predicted performance for job A) might not be adequate for predicted success on job A, while his second best score (predicted performance on job B) might be well above the value needed to predict success on job B.

2. Place each person on that job where his predicted score is farthest above the minimum score necessary to be considered satisfactory. This method is more

concerned with the total efficiency of the system rather than the extent to which each person is placed on the job he can perform best. It avoids putting anyone on a job where his performance will be substandard.

MULTIPLE CUTOFF SYSTEM

It was clearly pointed out in the discussion of the multiple regression system that the model used assumes linear relationships between the predictors and the criterion.[5] Such a system is often objected to on the basis that while for many traits there may be a linear relationship between predictor and criterion over most of the range, there may be some *minimum acceptable amount* of this trait that is necessary in order to be a successful worker. This kind of relationship between job performance and test is shown in Figure 3.7.

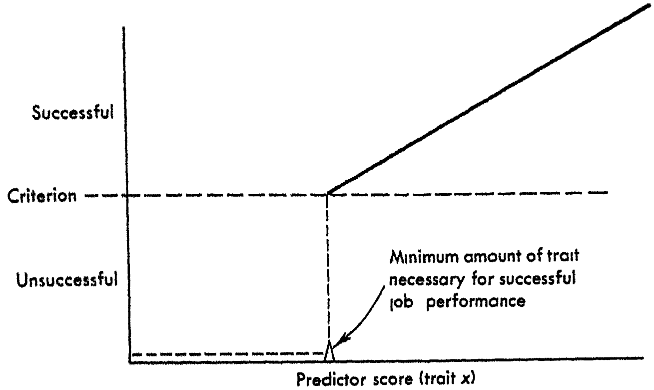


Figure 3.7. An illustration of the distortion of the predictor-criterion relationship that occurs if everyone below a certain criterion value is arbitrarily assumed unsuccessful, regardless of his other abilities.

The predictor-criterion function in Figure 3.7 shows what happens when one assumes that (1) there is some minimum amount of the predictor ability (trait X) necessary for job success, and (2) any lack or deficiency in trait X below this minimum *cannot* be compensated for by having a great deal of some other ability which has also been shown to predict job success. One example of such a situation might be an assembly job requiring both good vision and manual dexterity. Generally speaking, one might find that the better a worker's vision and the better his dexterity, the more successful that worker would tend to be on the job. However, there might be a point along the vision dimension beyond which no amount of dexterity would help.

The selection and placement procedure which takes into account this problem of minimum acceptable values is called the *multiple cutoff method,* meaning that a cutoff point is established separately for each predictor. Unless a person has a score above the cutoff on *all* predictors for a given job, he will not be placed on that job. Thus, no concept of additivity of traits exists with this method. Falling below minimum on any predictor will disqualify the individual.

Figures 3.8 and 3.9 show the regions of acceptance and rejection using the

[5] It also assumes linearity between predictors when expressed as in the equation $\hat{y}=a+b_1x_1+b_2x_2$.

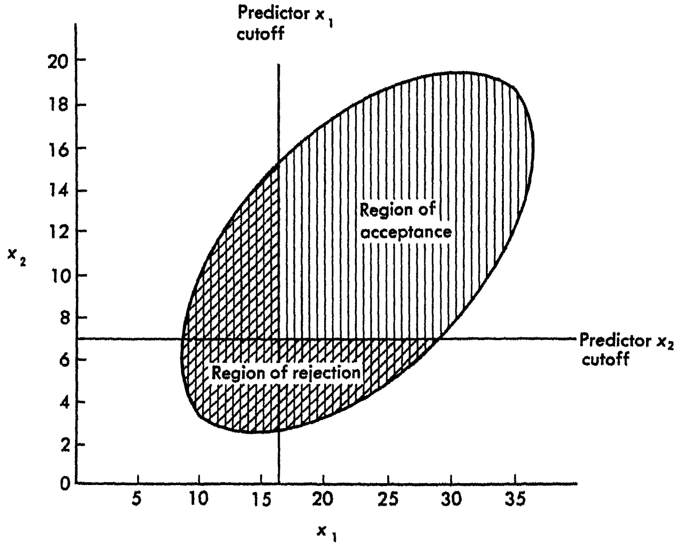


Figure 3.8. Regions of rejection and acceptance for job applicants formed by a multiple cutting score system situation where the cutting scores for x_1 and x_2 are 17 and 7 respectively.

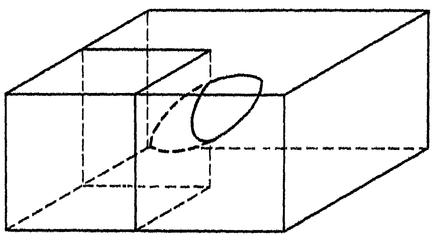


Figure 3.9. Extension of Figure 3.8 to three dimensions so that criterion scores for all individuals are represented. That section of the football scatterplot inside the small box is the rejects; that which protrudes outside is the acceptances.

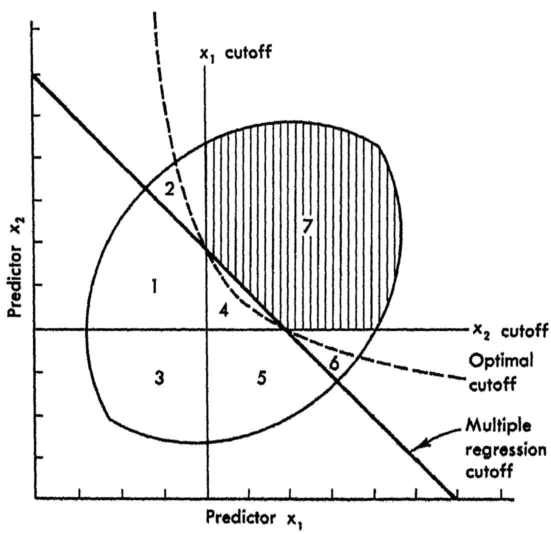


Figure 3.10. Comparison of the regions of acceptance and rejection for the multiple regression and multiple cutoff procedures. (See text for explanation.)

multiple cutoff system for data similar to that used to illustrate the multiple regression system in Figures 3.5 and 3.6.

Perhaps the best way to compare the two methods is to indicate how they differ in terms of who will be selected for the job. Figure 3.10 shows the cutoff lines for both selection methods. Note, first of all, that regardless of the method used, those people in area 7 will *always* be accepted and those people in areas 1, 3, and 5 will *always* be rejected. The people who will be differentially treated as a function of the selection procedure are those in areas 2, 4, and 6

Using the multiple regression selection system, all people in areas 2 and 6 will be accepted while those in area 4 will be rejected. The reverse will occur using the multiple cutoff procedure; the people in area 4 will be accepted and those in areas 2 and 6 will be rejected. Thus the question resolves to one of the relative desirability of these two groups of individuals.[6] The solution is mathematically complex and has been shown by Lord (1963) to be primarily a function of the reliability of the two predictors In fact, under most conditions probably *neither* procedure gives exactly the best solution in selecting that group of employees with the highest average criterion score. Instead, the optimal selection strategy appears to be some form of compromise between the two methods (see the dotted line in Figure 3.10).

DETERMINING CUTTING SCORES If one adopts the multiple cutting score technique, it becomes necessary to decide upon the separate minimum acceptable scores for each of the predictors. This is not an easy task since there is no specified "correct" way of setting up a score below which all people will be disqualified If the reader will refer back to Chapter 2 and reconsider the relationships involved in the selection ratio and the percent of employees considered satisfactory (the cutting score), he will begin to see how complex the problem is when two predictors are involved. Generally, the process of setting up cutting score values becomes one of trial and error in which different values for each predictor are tried. For each pair of cutting scores, the researcher must determine how high the average or composite criterion score of those selected is with respect to other cutting score combinations. He must also take into account the number of job openings with respect to the total number of applicants (the selection ratio measure).

ASSUMPTIONS, ADVANTAGES, AND DISADVANTAGES OF MULTIPLE CUTTING SCORES To summarize the points stated above, the method of cutting scores really assumes a nonlinear relationship between predictors and criterion. Second, it disavows the concept of substitution of test scores, at least in certain parts of the range. The one clear advantage is that it is usually an easy method for the personnel man to implement because no elaborate computational procedures or formulas are required. However, as was mentioned, a certain amount of trial and error is necessary to get cutting scores which will work in the most satisfactory manner. One of its more critical disadvantages is that it does not provide a single score for each individual which can be used to predict how successful he will be on one job relative to his success on another job. Thus, actual job placement via cutting scores can become exceedingly cumbersome.

[6] The reader is cautioned not to interpret equal areas of Figure 3.10 as representing equal numbers of employees. Because the scatterplot is really a *bivariate normal distribution* (i.e., a composite of two normals for a given correlation) the density (number of people) on the periphery of the scatterplot is much less than the number in the center. Therefore, area 4 probably contains many more people than areas 2 and 6 combined, even though the actual size of the areas would indicate the reverse.

PROFILE MATCHING SYSTEM

A third approach to employee selection and placement is the profile matching system. There are numerous versions of this method which differ primarily in terms of the way profiles are matched. However, the remaining aspects of the procedure are rather invariant from version to version.

The method itself is rather simple. If one has k variables (predictors) which are accepted as important to success on the job, then one measures all "successful" employees on the job on each of these k predictors. The scores are then averaged to obtain a "typical" profile of a successful worker. A hypothetical typical profile is shown in Figure 3.11. In this example ten predictors have been used to describe

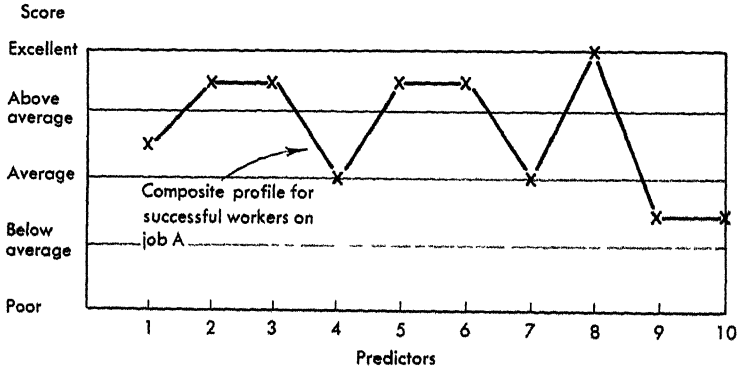


Figure 3.11. Example of a hypothetical profile composed of the average score for all workers considered successful on job A for each of ten predictors. Thus predictor 1 might be intelligence, predictor 2 might be clerical ability, etc.

the typical successful worker on job A. As the data indicate, a successful worker on job A will tend to have high scores (relative to other workers) on variables 2, 3, 5, 6, and 8. His scores on variables 1, 4, 7, 9, and 10 do not differ much from the average performance of workers in general. Once this sort of *ideal* profile has been obtained, it is then used as a standard against which the individual profiles of all new applicants are compared.

At this point two rather important questions arise in the profile method. First, how does one decide what predictors are relevant, that is, which ones should be included on the profile? Second, given that the profile elements have been successfully selected, how does one adequately judge the degree to which any individual's profile matches the ideal profile? The manner in which these two problems are solved can tremendously influence the eventual soundness and validity of any profile matching system.

SELECTION OF PROFILE ELEMENTS Each profile element is used as a predictor of job success, as are the predictors in the methods previously discussed. It is then certainly essential to establish the validity of each profile element prior to using it as a means of selecting and/or placing individuals on the job. What assurance do we have, for example, that poor or unsatisfactory workers do not have a composite profile that looks exactly like that shown in Figure 3.11? We have none at all, really, unless we proceed to find out empirically what the unsatisfactory com-

posite does look like by actually measuring a group of these people on the same traits and computing group averages.

It should be obvious that only those predictors which demonstrate a significant difference in mean scores between the satisfactory and unsatisfactory groups should be included on the ideal profile. Any trait which does not clearly differentiate between "good" and "poor" employees will only add error and confusion by being interjected into the selection process. Since validation of each trait is a necessary (but all too often ignored) step in profile item selection, it might be a legitimate question to ask why not simply use all profile predictors in a multiple regression equation (or perhaps even a multiple cutoff). Actually, the answer to this depends upon what method is employed for comparing profiles, as will be seen in the following section.

METHODS OF COMPARING PROFILES There are two quite different procedures which can be adopted in comparing the profiles of each individual to the ideal profile. One method selects those people who have profiles that *match* the composite most closely. This in turn results in a choice of procedures, depending upon how the term *match* is defined.

One way of defining a good match is to say that the closer the points of one profile are to the points of the other profile, the better the match. This method, then, uses the differences between the two scores on each trait to obtain a measure of similarity (or dissimilarity). The most usual procedure computes these differences, squares them, and then adds them up to get a similarity measure. Thus, if we have a profile with k traits, and if we further define

$$X_{ij} = \text{Score of person } i \text{ on trait } j$$
$$X_{sj} = \text{Score of standard profile on trait } j$$

then
$$D^2 = (X_{ij} - X_{sj})^2$$

and $\sum^{k} D^2$ would represent the degree to which the profile of person i matched the standard profile. The larger ΣD^2, the poorer the match. It is important to realize that the D^2 method does not concern itself at all with whether or not the scores of person i fall above or below the composite, that is, direction is *not* important with this matching procedure. All that counts is the closeness of profile points.

A second method of defining profile similarity is expressed in terms of our old friend the correlation coefficient. A high correlation between the scores of individual i's profile and the scores of the ideal profile indicates that the two profiles have similar *patterns*, i.e., individual i scores high on those traits on which the ideal profile also has high scores and he scores low on those traits where the ideal profile also has low scores. Figure 3.12 shows examples of profiles which illustrate how the use of different methods for assessing similarity can result in different individuals being selected for the job. Examination of Figure 3.12 quickly reveals that the general *pattern* of person B's scores duplicates that of the ideal or standard profile much more closely than do the scores of person A. However, the actual scores obtained by person A appear to be *closer* in value to those of the standard profile than the scores on individual B's profile. We can, therefore, hypothesize that individual A ought to have the lower (more desirable) ΣD^2 score while individual B should have the higher correlation (more desirable) with the standard. As the data in Table 3.3 indicate, such indeed turns out to be the case. When the values given in Figure 3.12 are used to compute ΣD^2, person A's *score* (ΣD^2_{as}) is 500, while

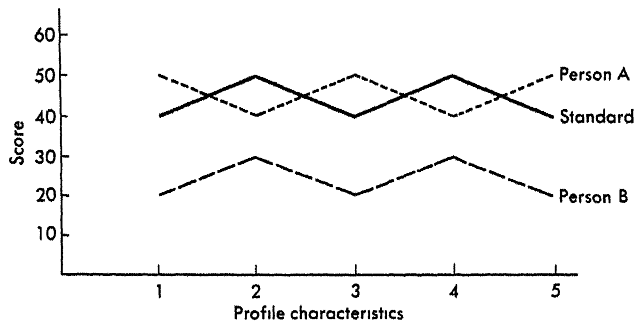


Figure 3.12. Example of effect of using different indices for profile matching. D^2 procedure would select person A, while correlation procedure would select person B

person B's score (ΣD^2_{bs}) is much larger, having a value of 2000. On the other hand, when correlations between profiles are computed, the correlation between profile A and the standard profile is computed $r_{as} = -1.00$, while the correlation between profile B and the standard, r_{bs}, turns out to be 1.00—complete agreement. Thus, if the D^2 method were used as a selection criterion, we would select person A; if we were to use correlation between profiles as a method, we would select person B.[7]

CHOOSING A PROCEDURE Which procedure is best is a question that can only be answered by empirical means in a particular setting. In all likelihood, however, neither the D^2 nor the correlation method is the best technique. If the traits on the profile have been selected on the basis of having significantly discriminated between good and poor employees (as indeed they most certainly ought to be selected), then the logical deduction is that high scores on a trait are to be desired and low scorers are to be avoided (or vice-versa, depending on the trait). If we assume, as we have generally, that the significant relationship between each trait on the

TABLE 3.3 *D^2 Values and Correlations Between Profiles of Individuals A and B and Standard Profile Shown in Figure 3.12*

	Score						
Item	*Person A*	*Person B*	*Standard*	D_{A-S}	D_{B-S}	D^2_{A-S}	D^2_{B-S}
1	50	20	40	10	− 20	100	400
2	40	30	50	− 10	− 20	100	400
3	50	20	40	10	− 20	100	400
4	40	30	50	− 10	− 20	100	400
5	50	20	40	10	− 20	100	400
	230	120	220	10	− 100	500	2000

$r_{as} = -1.00$ $r_{bs} = 1.00$

[7] It is beyond the scope of this book to explore the problems of profile analysis in more than a cursory way. It is hoped that the reader will become aware of some of the major difficulties which arise when we attempt to classify or select applicants using profiles, and that he may be interested in exploring the topic further via some of the references suggested at the end of the chapter.

profile and job success is positive and linear, then we would want to select people according to one of the following procedures:

1. Select those people whose profile points tend to be the highest, i.e., their *average* profile score is used as a selection index. Using this procedure a person could have a large ΣD^2 score and still be selected, as long as his profile points tended to be above the corresponding profile points for the standard. This procedure is equivalent to using a multiple regression selection model where each profile trait is a predictor and the regression weights are assumed equal for each predictor.[8] Low profile scores on one trait can be compensated for by high profile scores on another trait.

2. Select those people who have profiles with the highest average profile score and whose points *all lie above* their corresponding ideal profile counterparts. This, of course, is equivalent to a combination of the multiple cutoff selection method and the multiple regression method. The ideal profile points are used to establish minimum acceptable score values. All people thus qualifying are then evaluated via the multiple regression system. Such a procedure can probably only work in cases where the selection ratio is sufficiently small to enable one to employ rather stringent cutoff values. Certainly to use the average score on each trait for a group of successful employees as minimum acceptable values is creating a stiff hurdle for new applicants.

Either of these latter procedures seems a somewhat more justifiable way to use profiles for selection than the first two procedures, D^2 or r. The concept of an "ideal" profile in which deviations in any direction are considered "bad" can be seriously questioned on logical grounds.

MULTIPLE HURDLE SYSTEM

Most selection situations involve attempts to predict later success on some task through the use of one or more predictive measures obtained at the time of job application. However, some selection situations such as management training include somewhat lengthy periods and ultimate evaluation after quite some time, but with interim evaluations or hurdles at various points of progress.

Consider the situation illustrated in Figure 3.13. Here we have diagramed a training program that might be used by a large corporation as a means of screening, training, and placing new college graduates within the corporation. The company initially hires a given number of college graduates, perhaps using college grades, interviews, letters of recommendation, and tests as a means of selecting people. All hires are told that their selection is on a probationary basis and that they will be continually evaluated during their training program. If performance during training is not satisfactory, they may be released from the program.

It is certainly in the company's interest to make an accurate decision about each individual as early as possible. Similarly, it is equally in the best interest of the employee that a decision be made as early as possible. However, the degree to which it is possible to predict success as a result of the training program increases in correctness (that is, validity increases) the longer we are able to observe the performance of the individual during training. By the end of the third evaluation period we should certainly be able to predict much more accurately whether a trainee is going to finish the course successfully than we were able to do at the time he was hired. The situation is quite analogous to the problem of predicting the final

[8] It would also be necessary that each predictor have an equal variance.

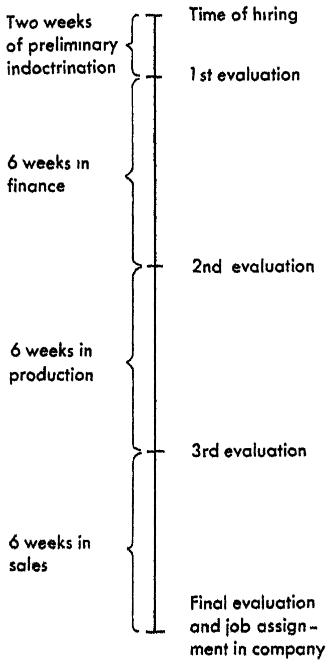


Figure 3.13. Example of a typical management training program used to train, evaluate, and place newly hired college graduates.

grades of college students. Obviously, one can make better predictions by the time the student is starting his senior year than at the time he enters college. Figure 3.14 illustrates the change in validity that one might logically expect to occur in a situation such as that diagramed in Figure 3.13.

In one sense the mechanics of a situation as shown in Figure 3.13 are identical to the more usual multiple predictor situations: A number of predictors of success are available, but to obtain each additional predictor it is necessary that additional time and money be invested into that trainee. Sequential predictors are used in several ways. Most frequently, one of the following methods is employed.

1. A person must score above some minimum desirable score at each evaluation stage. Thus, each stage becomes a hurdle which the trainee must clear if he is to be kept in the program.

2. A composite multiple regression is computed at each successive evaluation point, and the probability of success is computed for each person remaining in the program. Whenever this probability drops below some arbitrary value (for example, 25 percent), he is dropped from the program.

PROBLEM OF RESTRICTION OF RANGE One difficulty that emerges in sequential selection situations is a problem known as the effect of "restriction of range" upon validity estimates. If we have used predictor 1 to select people initially, and then if we subsequently compute the correlation between predictor 1 and the criterion or compute the correlation between some other predictor 2 and the criterion, our computed validity coefficients r_{1c} or r_{2c} are likely to be underestimates of the validity we would have obtained had *no* preselection taken place. By preselecting we have restricted the range of ability (and therefore predictor scores) which will reduce

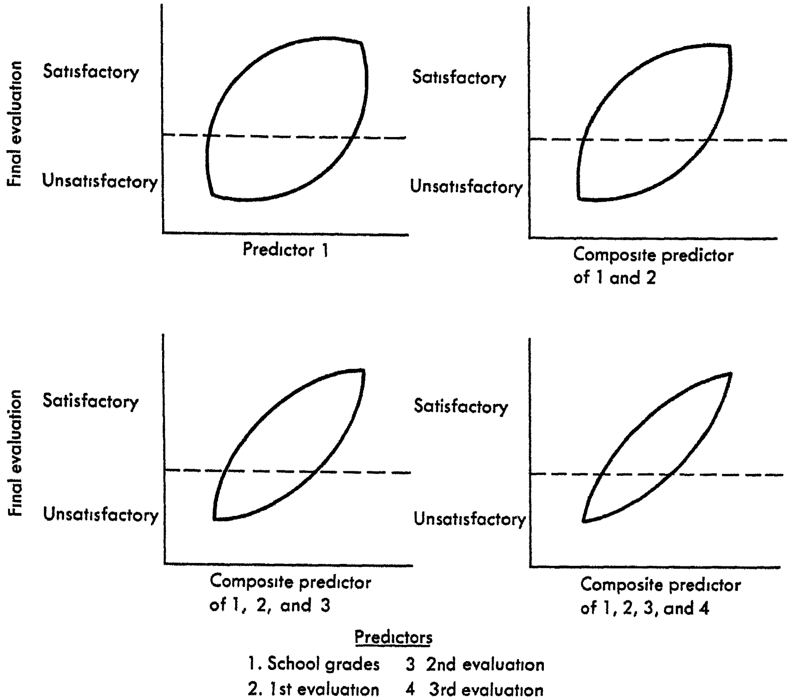


Figure 3.14. Diagrams showing the increase in validity to be expected as one obtains additional information about trainees as they progress through the different training phases shown in Figure 3.13.

the correlation coefficient. Indeed, our predictor 1 acts in a manner similar to the control variable in partial correlation; since it has already accounted for a portion of the variance, the correlation r_{2c} will be reduced. To get an estimate of what the validity R_{2c} truly is, one can use the correction formula.[9]

$$R_{2c} = \frac{r_{2c} + r_{12} r_{1c}\left(\frac{S_1^2}{s_1^2} - 1\right)}{\sqrt{\left[1 + r_{12}^2\left(\frac{S_1^2}{s_1^2} - 1\right)\right]\left[1 + r_{1c}^2\left(\frac{S_1^2}{s_1^2} - 1\right)\right]}} \tag{3.8}$$

where R_{2c} = corrected validity of predictor 2 for entire group of applicants
r_{2c} = computed validity for predictor 2 based upon restricted group
r_{1c} = validity of initial screening instrument, predictor 1
r_{12} = correlation between predictors 1 and 2
S_1^2 = variance of predictor 1 in original group
s_1^2 = variance of predictor 1 in restricted group

[9] For a more extensive discussion of this topic and its importance to selection and placement see R. L. Thorndike. *Personnel selection.* Wiley, New York, 1949.

MODERATORS

One of the more important concepts in selection and placement theory is the concept of the *moderator variable.* Sometimes referred to as a population control variable,[10] a moderator variable may be viewed as *any variable which, when varied systematically, has an effect upon the magnitude of the relationship between two or more other variables.*

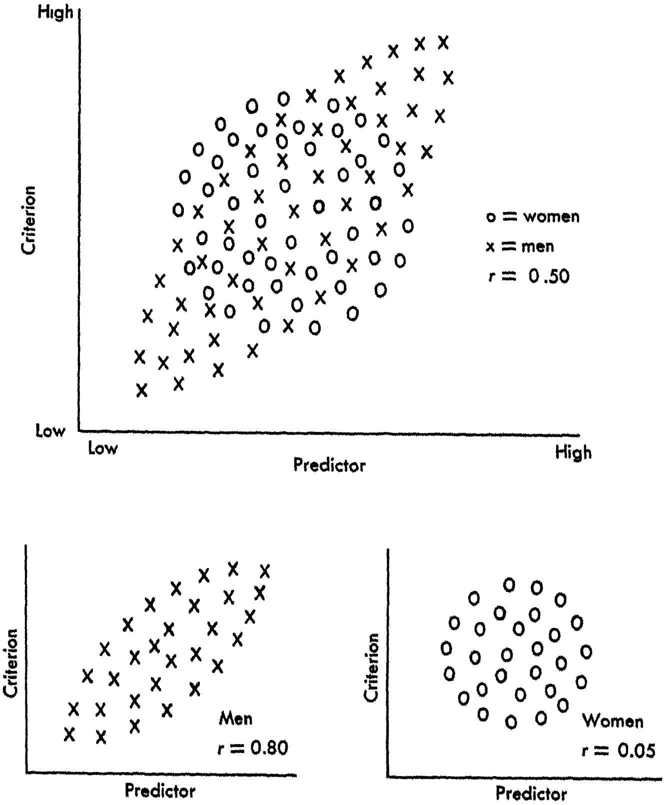


Figure 3.15. Diagrams illustrating the effect on validity which can occur through the utilization of a moderator variable such as sex.

Perhaps a hypothetical example (Figure 3.15) of how a moderator might function will serve to illustrate its influence upon the selection process. The top scatterplot illustrates a general validity of 0.50 between the predictor and a criterion. However, the "population" represented in the scatterplot is one which includes both sexes, viz., both men and women are grouped together in determining validity. Even a casual inspection of the top scatterplot indicates (if men and women are coded differently as has been done here) that the pattern of scores observed for men differs from that observed for women. To get a clearer picture of exactly how they

[10] R. H. Gaylord and J. B. Carroll. A general approach to the problem of the population control variable (abstract). *American Psychologist,* 1948, 3, 310.

differ, the two lower scatterplots in Figure 3.15 show the predictor-criterion relationships separately for men and for women. Now the difference is striking. For the men we observe a high positive relationship—one that produces a validity of 0.80. For the women, on the other hand, we see that there is virtually no relationship between the predictor and the criterion. The validity for women is 0.05.

The moderator variable in the above example is, of course, the variable of sex. The relationship between predictor and criterion is drastically affected by varying the moderator. The question "what is the validity of my predictor" clearly becomes more complex. What initially appeared to be a moderately respectable validity has now turned into two quite distinct and separate validities—one very high and one very low. One name for these latter validities might be *conditional validities*, that is, the validity of the predictor *given that the population consist of women* or *given that the population consists of men*.[11] An interesting characteristic of moderator variables is that a moderator need not have any direct relationship with either the predictor or the criterion variable (that is, r_{ym} and $r_{xm} = 0$).

EXAMPLES OF MODERATORS

Actual examples of moderators have been found in a number of research investigations. Vroom (1960), for instance, found quite marked moderator effects using degree of motivation of managers and first-line supervisors as the moderating variable. All men studied were employees in either the Chicago or New York plant of a national delivery service company which specialized in delivering small packages and parcels from department and other retail stores to private residences Data from the study which best illustrate the moderator concept are given in Table 3.4.

TABLE 3.4 *Relationships Between a Measure of Nonverbal Reasoning and Ratings of Job Performance as a Function of the Motivation of the Worker*

		Correlations Between Nonverbal Reasoning Scores and Supervisory Ratings of			
	N	*Overall Performance*	*Summary Appraisal*	*Overall Results (Prod. Cost Qual.)*	*Overall Results (Skill, Methods, Effort)*
High motivation group	31	0.47**	0.56***	0.33**	0.21
Moderate motivation group	28	0.06	− 0.04	− 0 19	− 0.05
Low motivation group	32	− 0.07	− 0 23*	− 0.17	− 0.31**

* = Chance probability less than 5 times in 100
** = Chance probability less than 1 time in 100
*** = Chance probability less than 1 time in 1000
SOURCE: Adapted from Victor H. Vroom. *Some personality determinants of the effects of participation*. Prentice-Hall, Englewood Cliffs, N.J., © 1960, p. 55.

[11] The reader who has had some exposure to the Analysis of Variance and multiple regression analysis in statistics may recognize that the presence of a moderator would be indicated by an interaction between the moderator and predictor in the former, or by a significant cross-product term in the multiple regression equation.

All supervisors were divided into three groups based upon their assessed degree of motivation using a composite of several motivation indices obtained in the research. Validities for a test of nonverbal reasoning ability were then obtained for each of four different types of supervisory ratings of these men. This was done separately at each motivation level. As Table 3.4 shows, the test was apparently a quite valid predictor of how high a man would be rated by his supervisor *if only men with high motivation were considered.* If we systematically vary motivation by moving down to the groups having only moderate or low motivation levels, we see a corresponding systematic change in the relationship between the test and the criterion. The lower the motivation of the employee, the less the validity of the predictor; in fact, the validities even become negative for the low motivation groups.

Other examples of moderators can be found in studies by Dunnette and Kirchner (1960) and Ghiselli and his coworkers (1956, 1960). The work of Dunnette and Kirchner has been directed primarily at identifying job-related moderators, that is grouping people into jobs which are similar in terms of their responsibilities to get maximum prediction within each job group. Ghiselli's method might be called a "variable-free" moderator system: People are grouped simply on the basis of how well their success can be predicted with no direct reference to any external variable. Fredericksen and Gilbert (1960) have also done research on moderators to determine the degree to which a moderator's effect is likely to be consistent over time. They found that a moderator identified in a 1954 study (Fredericksen and Melville, 1954) was still operating in a 1960 follow-up.

MODERN VERSUS TRADITIONAL SELECTION THEORY

The concept of the moderator variable perhaps best illustrates the modern trend in selection and placement emphasis. Traditionally, selection and validation have been problems which were viewed as being best solved by simply establishing a criterion which appeared to be reliable and a predictor which could best predict that criterion. The emphasis was almost completely upon the establishment of a high validity with little or no thought toward exploring the many additional variables which, when varied, might add to or subtract from the obtained correlation. The general motto which all too often seemed to typify selection methodology was the slogan "If it works, use it!"

Without question, this policy was responsible for quite different developments in industrial psychology. First, it probably contributed to the degree to which psychologists were accepted into industry. Management is generally oriented toward positive results as represented by improved selection, and is not overly concerned with how it is accomplished. Unfortunately, however, this orientation is also probably responsible for the fact that validities in prediction have not risen substantially (if at all) during the past 50 years—a rather disturbing commentary upon the efforts of psychologists engaged in this type of work. In a 1955 review of a large number of validity studies, Ghiselli (1955) indicated that it is indeed an unusual event to obtain a validity coefficient of 0.50 or better. Figure 3.16 presents frequency distributions presented bv Ghiselli of validity coefficients of varying magnitudes for different types of jobs. Note that only in the distribution of validities for clerical workers using intelligence tests as predictors and proficiency measures as criteria are there a large number of validities above 0.50.

The current interest in moderators is representative of a broader and somewhat

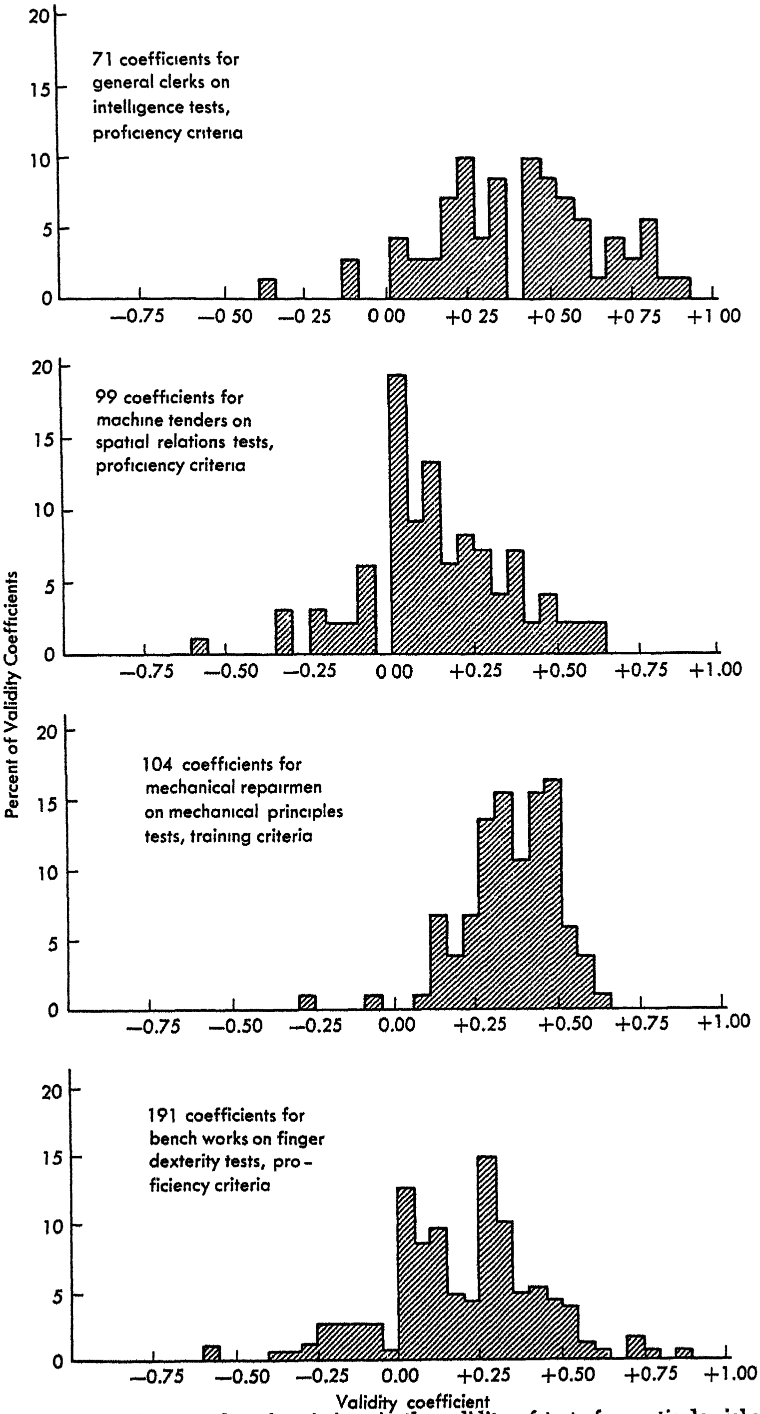


Figure 3.16. Examples of variations in the validity of tests for particular jobs. (From E. E. Ghiselli. *The measurement of occupational aptitude.* University of California Press, Berkeley, 1955.)

more sophisticated approach toward selection. It can be traced about to when Toops (1948) made an appeal for psychologists to consider the possibility that by stratifying people (for example, workers) systematically according to personal variables, one should be able to improve prediction. His method of classification, which he referred to as the *addend procedure*, is the forerunner of moderators.

DUNNETTE'S SELECTION MODEL

Perhaps the current view toward selection methodology can best be represented by the selection model proposed by Dunnette (1963). This model is shown in the diagram presented in Figure 3.17 and is designed to point out the maze of complexities and interrelationships which exist in the selection situation. The model may be viewed as more than an attempt to merely point out the dynamic nature

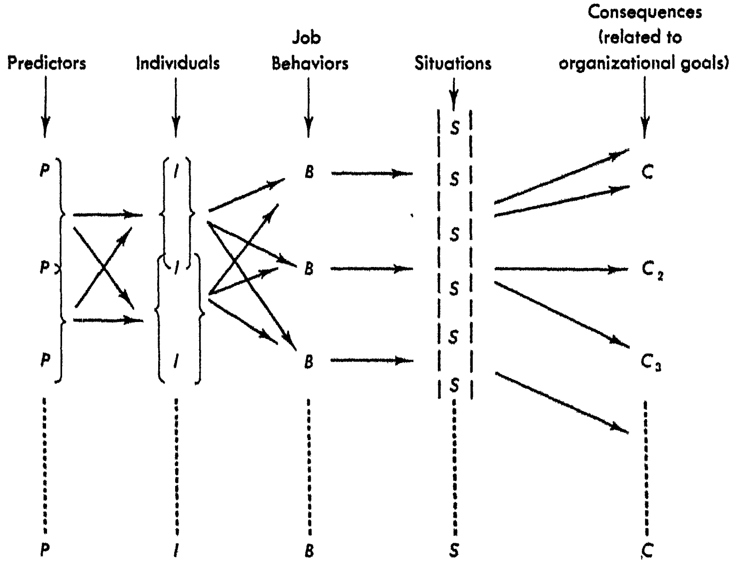


Figure 3.17. Modification of Dunnette's selection model, a diagram showing the complexities of any selection system. Every validation is subject to moderating effects of predictors, individuals, and situations. (Adapted from M. D. Dunnette. A modified model for selection research. *Journal of Applied Psychology*, 1963, *47*, 317–323.)

of selection—it also represents a plea for psychologists to take advantage of these dynamics and use them to best advantage in order to improve predictability.

One can probably understand the point of view represented by the model in terms of the exact description used by Dunnette (1963, p. 318):

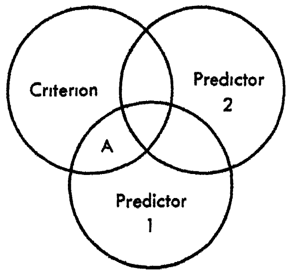
> Note that the modified prediction model takes account of the complex interactions which may occur between predictors and various predictor combinations, different groups (or types) of individuals, different behaviors on the job, and the consequences of these behaviors relative to the goals of the organization. The model permits the possibility of predictors being differentially useful for predicting the behaviors of different subsets of

individuals. Further, it shows that similar job behaviors may be predictable by quite different patterns of interaction between groupings of predictors and individuals or even that the same level of performance on predictors can lead to substantially different patterns of job behavior for different individuals Finally, the model recognizes the annoying reality that the same or similar job behaviors can, after passing through the situational filter, lead to quite different organizational consequences.

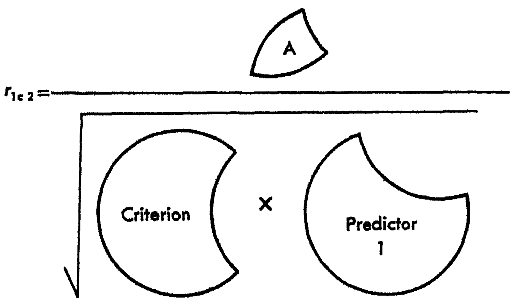
The current trend in selection represented by the awareness of moderators and by Dunnette's selection model should result in progress in both the increased efficiency of selection and the degree of understanding of the dynamics of accurate prediction.

SUPPRESSOR VARIABLES

No discussion of selection would be complete without some mention of *suppressor variables*. In one sense a suppressor variable is similar to a moderator variable in that it is defined as "a variable which can have an effect upon the magnitude of a given predictor-criterion relationship even though it has little or no relationship to the criterion variable itself." The dynamics of a suppressor variable in prediction can best be understood by reviewing again the concept of a partial correlation and its related measure, the *semipartial* correlation. In Chapter 2 it was stated that if one had two predictors and a criterion that were intercorrelated as shown here

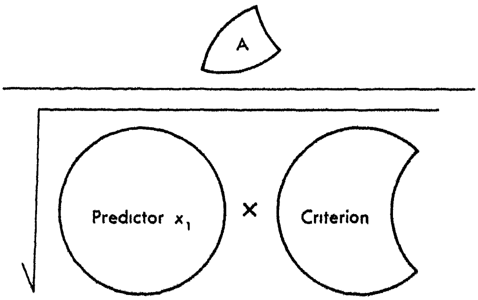


then the partial correlation between the criterion and predictor x, which is r_{1c2}, was defined as the correlation between x_1 and C after the effects of x_2 have been partialled out of both, so

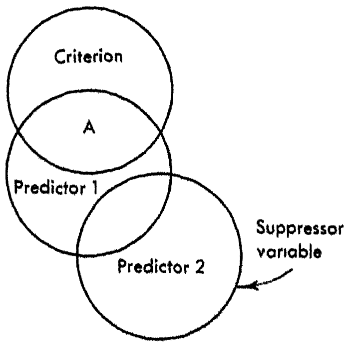


Suppose we only want to remove the effects of x_2 from the criterion prior to computing the correlation. Such a correlation is called a semipartial or part correlation. For example, we might be interested in the correlation between intelligence test scores (our predictor x_1) and final skill level at the end of a typing training program (the criterion). x_2 might represent the initial skill level of all employees in terms of their typing speed prior to taking the training course. Thus, we want to remove the effects of initial skill level upon final performance before computing the validity of our intelligence test. Our semipartial correlation now becomes

$r_{1(c\ 2)} =$

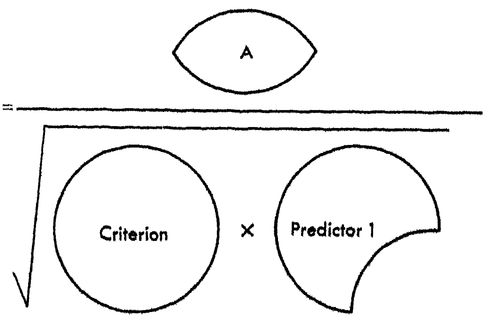


The mechanism of a suppressor variable is identical to that shown above except (1) generally, variable x_2 has only a slight (if any) relationship to the criterion, and (2) one is interested in removing its effects from predictor x_1. The general situation can therefore be diagramed as



and the "suppressed" correlation is

$r_{c(1,2)} =$



One cannot predict with complete certainty whether partial or semipartial correlations will be larger or smaller than the simple correlation existing between the variables, since the size of both the numerator and the denominator is affected by the partialling process. The only time this is *not* so is when the variable being partialled out is only related to one of the two other variables, as in the case of the suppressor. In such a situation only the denominator is subsequently affected (variance is removed) and the resulting semipartial correlation is larger than the simple unpartialled correlation between variables.

CROSS-VALIDATION

One feature of most multiple prediction selection systems is that in their development one typically tends to capitalize upon the chance variation that exists in the sample of employees being used for purposes of validation. This is particularly true with the multiple regression model, but applies as well to the multiple cutoff procedure. Because the multiple regression model has least-square properties, i.e., we deliberately minimize the errors in predicting *our particular sample,* it is likely that if we now apply our equation to a new sample (from the same population) we will not find our prediction as efficient as before. Thus, our computed R^2 is an overestimate of what the future validity of our prediction system is apt to be, since using our equation for purposes of prediction automatically implies applying it to new samples of workers. This expected drop in R^2 is known in statistics as the *shrinkage problem* and can best be illustrated by examining Figure 3.18.

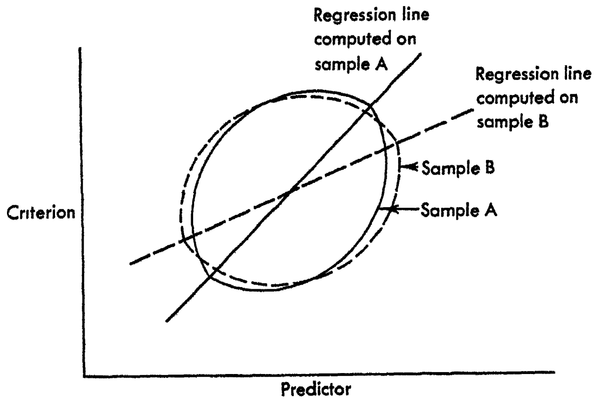


Figure 3.18. Regression lines of best fit for two different samples of workers belonging to the same overall "population."

In Figure 3.18 we have two samples of individuals. Each represents a random sample drawn from or belonging to the same population. For example, sample A might represent all job applicants for job X during the odd-numbered months, and sample B might represent all job applicants during the even-numbered months for a particular year. It would be highly unusual, even with very large numbers of applicants in each sample, for the two samples to be identical in terms of their scatterplots. Since their scatterplots can be expected to vary due to sampling error,

the correlation between the predictor and criterion (validity) can also be expected to vary somewhat, as can the regression equation computed on each sample.

Suppose we took the regression equation computed on sample A and used it to predict scores from sample B. We obviously could not do as good a job in minimizing Σd^2 using the A line with sample B as we could using the B regression line—after all, the B line by definition minimizes Σd^2 *for that sample*. Any other line will therefore have a larger error associated with it. Thus R^2 must be reduced correspondingly.

There are formulas available for estimating the amount of shrinkage one can expect when using this equation on a new sample. One such formula is

$$R_s^2 = 1 - \left[(1 - R^2) \, \frac{n-1}{n-k-1} \right] \tag{3.9}$$

where R_s^2 = shrunken multiple correlation squared
R^2 = multiple correlation squared obtained from validation sample
n = number of people in validation sample
k = number of predictors in regression equation

It is best, however, to cross-validate the equation by obtaining a second sample and trying it out to see how well it does predict. If there appears to be a very large drop, one may want to revise the equation (perhaps by combining both samples in one group). Large shrinkage is most often found when the sample sizes are small and/or the number of predictors is large relative to the sample size.

Mosier (1951) has discussed a number of *types* of cross-validation that can be conducted depending upon the design of the study and whether one is concerned about generalizing only to a new sample or if broader generalizations concerning the prediction equation are desired (for example, to different sexes, different criteria, etc.). The former is called a case of *validity generalization*; the latter is a case of *validity extension*. Of course, greater shrinkage would be expected in the latter case, and formula 3.9 only applies for cases of validity generalization.

REFERENCES

Catell, R. B. r_p and other coefficients of pattern similarity. *Psychometrika*, 1949, *14*, 279–298.

Cronbach, L. J., and G. C. Gleser. Assessing similarity between profiles. *Psychological Bulletin*, 1953, *50*, 456–473.

Dunnette, M. D. A modified model for test validation and selection research. *Journal of Applied Psychology*, 1963, *47*, 317–323.

Dunnette, M. D., and W. K. Kirchner. Psychological test differences between industrial salesmen and retail salesmen. *Journal of Applied Psychology*, 1960, *44*, 121–125.

Fredericksen, N., and A. C. Gilbert. Replication of a study of differential validity. *Educational and Psychological Measurement*, 1960, *20*, 759–767.

Fredericksen, N., and S. D. Melville. Differential predictability in the use of test scores. *Educational and Psychological Measurement*, 1954, *14*, 647–656.

Ghiselli, E E. *The measurement of occupational aptitude.* University of California Press, Berkeley, 1955.
Ghiselli, E. E. Differentiation of individuals in terms of their predictability. *Journal of Applied Psychology,* 1956, *50,* 374–378
Ghiselli, E. E The prediction of predictability *Educational and Psychological Measurement,* 1960, *20,* 3–8.
Horst, P. *The logic of personnel selection.* University of Washington, Seattle, 1961, 66.
Lord, F. M. Cutting scores and errors of measurement. *Psychometrika,* 1962, *27,* 19–30.
Mosier, C. I. Symposium· The need and means of cross-validation I. Problems and designs of cross-validation. *Educational and Psychological Measurement,* 1951, *11,* 5–11.
Nunnally, J. The analysis of profile data *Psychological Bulletin,* 1962, *59,* 311–319.
Sawrey, W. L., L Keller, and J. J Conger. An objective method of grouping profiles by distance function and its relationship to factor analysis. *Educational and Psychological Measurement,* 1960, *20,* 651–674.
Toops, H. A. The use of addends in experimental control, social census, and managerial research. *Psychological Bulletin,* 1948, *45,* 41–74.

4 TESTING IN INDUSTRY

Industry generally recognizes that psychological testing is the function of the trained psychologist. Whereas many other areas of work such as incentive development, training, accident reduction, market research, and advertising effectiveness are studied by professionals from other disciplines, testing is most often reserved for the psychologist.

As stated in Chapter 1, tests were used in industry prior to 1910 by Hugo Münsterberg in connection with various problems in his research for the Boston Elevated Railway Company. Testing in industry has steadily increased since that time. During the two world wars, emphasis was placed on psychological tests, and the significant contributions made resulted in a tremendous impetus to the field during and immediately following these periods. Since 1945 the use of tests has become commonplace. In fact, as we will discuss a little later, the problem is often more of restricting the use of tests than of trying to promote their acceptance.

This chapter is intended to introduce the reader to the use of tests in industry by examining some of the advantages and disadvantages of psychological tests, by examining various types of tests which are currently in vogue, and by looking at some of the special problems which accompany the use of tests. Since testing is so widespread and the number of different tests used are so numerous, it should be pointed out that this chapter only introduces the student to the world of tests.

CHARACTERISTICS OF PSYCHOLOGICAL TESTS

Merely assembling a batch of questions does not produce a psychological test. A test may best be defined as measuring a *standardized sample of human behavior*. As such, it must meet certain basic requirements. First, the sample of behavior should be both large enough and representative enough of the class of behavior we are measuring to allow us to generalize and predict from the test results. Second, by requiring the test to be *standardized*, we mean that the series of questions or tasks should be uniformly administered whenever subjects are given the test. Finally, all psychological tests must have the prerequisite characteristics of validity, reliability, and norms.

Throwing together a series of questions regardless of whether they concern psychological topics does not elevate them to the level of psychological testing. For example, many newspaper columns and popular books of the parlor-game variety

present a series of questions; the person scores the answers and rates himself as "excellent," "good," "bad," or "awful." These are not psychological tests; they are merely batches of questions. Of course, an individual often has an overwhelming curiosity to find out something secretly, and so the self-testing plan is very popular. The subject matter of these tests usually has wide appeal. For example, a column may ask you to determine whether you are an outstanding husband by answering a few questions. A person may have inscrutable wisdom, but the only way he can develop a *test* to measure such qualities is to have a series of questions that have been answered in one way by successful husbands (whoever they are) and in a significantly different way by unsuccessful husbands.

TEST NORMS

Since the topics of reliability and validity were covered in detail in Chapter 2, no further explanation of their characteristics seems necessary here. However, the topic of test *norms* is one of considerable importance and needs additional clarification. Norms should be developed as a source of reference on all tests used in selection.

A norm is a standard of reference; it enables one to understand the meaning of a test score. Depending on the test, a raw score may be reported in various ways: Total time to complete the test, number of items correct, or number of items attempted are a few of the raw scores obtained on tests. Actually, the raw score by itself on a psychological test is usually a meaningless figure. For example, on one test a score of 240 seconds may be poor, whereas on another test a score of 75 seconds may be exceptional. The problem is further complicated when a score of 180 seconds must be compared with a score of 95 items correct. Without the use of a norm, such comparisons would be impossible; at best, they would resemble an attempt to compare apples and peaches.[1]

The two most widely used systems of norms are *percentiles* and *standard scores.* Both measures indirectly give information on the test performance of the individuals relative to a known population. They also show the relative position of one person in the group to the group as a whole. On a test to measure stenographic ability, an individual's raw score might be 105 words per minute with one error in transcription. Unless we had information about the range of speed based on many people, we could not know whether this rate of speed was good or poor. However, if this test has been given to 155 gainfully employed stenographers and this individual's score placed her in the 90th percentile, we would be in a position to estimate her ability to take shorthand as "exceptional" because she exceeds 9 out of 10 girls in shorthand speed. If this same individual types at a speed of 45 words per minute, and if this score is equivalent to the 20th percentile, we could conclude that 8 out of 10 employed stenographers are faster typists. Actually then, shorthand speed and typing speed can be compared. On the basis of the norms available, we can make a direct comparison of this girl's performance on these two tests, and our conclusion that she is very good in taking dictation but poor in typing is justified. Whether she is hired depends on the needs of the job; it is most probable that she would have to improve her typing speed before job placement would be possible.

[1] While norms are necessary to attach any real meaning to a test score, they are not necessary to the process of predicting some criterion of job success from the predictor scores.

For some clerical jobs speed is of great importance, but in others accuracy is more important. For example, speed rather than accuracy might be sought for addressing circulars. On the other hand, extreme accuracy would be necessary for filing important papers. Sometimes separate norms on speed and accuracy are necessary, as well as norms based upon a combination of the two measures.

In the establishment of norms, specific information should be available not only about the size of the group measured, but also about such facts as age level, whether the group is gainfully employed, whether it is a college population, and whether it consists of a normal unselected sampling. For example, a person who tests in the 70th percentile of a normal population on an intelligence test can be considered as having above average intelligence. However, his probable success in a Grade A college or graduate school would be questionable. In other words, a norm is meaningful only when the characteristics of the population upon which it is based are known.

TYPES OF TESTS

Psychological tests may be categorized into various groupings depending upon the particular purpose of the classification system. Before proceeding further, it might be valuable to consider some of these classification systems.

TYPE OF BEHAVIOR MEASURED

Probably the most frequently used classification system is based on the type of behavior that tests purport to measure. Thus we find tests designated as intelligence tests, personality tests, interest tests, vision tests, music tests, art tests, mechanical tests, verbal tests, etc., where each group is fairly behavior specific.

ACHIEVEMENT AND APTITUDE TESTS

Very frequently authors will distinguish between tests of *aptitude* and tests of *achievement*. The former is supposedly a measure of a person's potential in a given area, while the latter is a measure of a person's current skill or ability at the moment of testing. Since the same test can often be considered both an achievement test and an aptitude test depending upon use, this classification system is often a fuzzy one. Thus, with many tests one can (1) measure the amount of present skill, and (2) use the present score to predict future performance.

PAPER-AND-PENCIL TESTS AND PERFORMANCE TESTS

Many tests are of the *paper-and-pencil* variety: The testee simply receives a test paper or printed booklet containing the test questions, and he records his responses in some written manner on the answer sheet that is usually provided. Many other tests, however, do not require a written response—rather, they involve some sort of manipulation activity such as handling pegs or blocks, or assembling mechanical objects. These latter tests are referred to as *performance* tests.

SPEED AND POWER TESTS

Some tests are constructed so that every item is very easy—the task is to complete as many items as possible in a short time. When test performance is based primarily upon the speed with which one works, the test is referred to as a *speed* test. The other extreme would be a test where the items were difficult and the person was given as much time as necessary to complete the items. In such tests a person's score is based exclusively upon his ability to answer the questions correctly, no matter how long (within reason, of course) it takes. This type of test is called a *power* test.

INDIVIDUAL AND GROUP TESTS

There are a number of tests which are designed to be administered individually; that is, they cannot be given simultaneously to two or more people by a single examiner. An example would be the Stanford-Binet Intelligence Scale. Very often *individual* tests are used for clinical evaluation. *Group* tests are those which may be taken by many people at the same time. For industrial testing, group tests are generally preferred because they are more economical to administer.

LANGUAGE AND NONLANGUAGE TESTS

Sometimes it is important to distinguish between those tests which require a knowledge of a particular language (such as English) in order to understand either the test instructions or the test items themselves. All such tests are called *language* tests because performance on them depends partially upon the language ability of the testee, regardless of the type of ability the test is designed to measure. In some cases it is desirable or necessary to avoid the language bias of a test. For example, to test the mechanical ability of people who are illiterate using a test that has complicated written instructions would be quite inappropriate. To solve this problem, tests have been constructed in certain areas which are "language-free" tests: They require no language skill on the part of the testee. These are referred to as *nonlanguage* tests.

TESTS VERSUS OTHER SELECTION DEVICES

In this chapter our attention is devoted exclusively to examining psychological tests. However, tests are by no means the only predictive tool employed by the psychologist in a selection situation. Other standard selection aids are letters of recommendation, application blanks, and interviewing. Comprehensive selection programs will use these in addition to tests as part of the total selection process. In Chapter 5 these other methods will be discussed in greater detail.

PROS AND CONS OF TESTING

Psychological testing in industry should be approached with caution. Testing is often regarded as a fascinating subject by the novice. In many instances the desire to introduce a testing program is fostered in the hope that a solution will be

obtained to a "no solution" problem. Even when tests are used, it is advisable to conduct an interview, check the application form information, and use the various other techniques recommended by good personnel procedure.

The expected result of improving the selection of applicants is that training costs and labor turnover are decreased, production is increased, accidents and the probability of accidents are reduced, and morale may be increased. However, the mere introduction of a valid testing program cannot act like a magic wand to wave away all difficult problems. Testing is no panacea, although testing in industry can have value if used correctly. However, an improvement in job performances of 10 to 20 percent is often all that can be expected. Although better results are sometimes obtained, this is the exception rather than the rule.

Table 4.1 shows some data collected by Ghiselli (1955) on the efficiency of testing

TABLE 4.1 *Average Validity Coefficients for Various Types of Tests*

	Type of Criterion	
Type of Test	*Training*	*Job Proficiency*
Intellectual abilities		
Intelligence	38 *c*	19 *c*
Immediate memory	29 *c*	19 *c*
Substitution	26 *c*	21 *c*
Arithmetic	41 *c*	21 *c*
Spatial abilities		
Spatial relations	31 *c*	14 *c*
Location	24 *c*	15 *c*
Perception of details		
Number comparison	26 *c*	21 *c*
Name comparison	25 *c*	21 *c*
Cancellation	29 *c*	20 *c*
Pursuit	19 *c*	17 *c*
Perceptual speed	39 *c*	27 *c*
Mechanical comprehension		
Mechanical principles	34 *c*	26 *c*
Motor abilities		
Tracing	16 *c*	16 *c*
Tapping	12 *c*	14 *c*
Dotting	14 *c*	15 *c*
Finger dexterity	22 *c*	19 *c*
Hand dexterity	38 *a*	14 *c*
Arm dexterity	30 *b*	17 *c*
Personality traits		
Personality	16	21 *c*
Interest	14 *a*	27 *c*
Personal data	44 *c*	41 *c*

Less than 100 cases = no letter symbol
[a] = 100 to 499 cases
[b] = 500 to 999 cases
[c] = 1000 or more cases

SOURCE: E. E. Ghiselli. *The measurement of occupational aptitude.* University of California Press, Berkeley, 1955.

for various types of industrial tests. Ghiselli examined large numbers of validation studies and classified them in terms of the type of test being used and the type of criterion involved. All validities within a particular category were then averaged to give the figures shown in Table 4.1. Note that only three of the values in Table 4.1 exceed 0.40 in size, and that two variables are *not* tests but are personal history items. While the process of averaging used by Ghiselli definitely masks some very substantial validities, it should be sufficiently clear that testing is not a complete solution to the selection problem.

From the point of view of the statistician, such correlations are, of course, too low for individual prediction purposes. The relationship is nevertheless useful, for even such slight help in predicting success is better than if no tests were given, as was made clear in the earlier discussion of the selection ratio.

In the discussion of job satisfaction and morale in later chapters, we will find that many other factors, in addition to an individual's ability, help to determine performance. Compelling reasons for continuing on a job can compensate for an employee's limited ability. On the other hand, for many reasons a person may be a failure on a job even though he has the ability to be successful. Success or failure depends upon other things besides ability. Most industrial testing programs limit the functions to be measured to certain specific abilities and do not consider the other contributing factors. Furthermore, the claim cannot be made that even ability is perfectly measured.

It is no wonder that psychological tests in industry cannot do a perfect job at the present time. Anyone who expects perfection is unaware of the real situation. For example, a girl who worked in a certain watch factory had an abnormally poor score on a finger and tweezer dexterity test. In this instance better test scores were correlated with better job performance. This girl actually had a poor production record, and since salary was determined on a piece-rate basis, her earnings were very low. She was also rated as a poor worker by her supervisor. Nevertheless, she was among the workers who had the longest tenure of employment. An interview with her revealed that she was under extreme financial pressure and simply could not afford to be out of work one day. Moreover, experience had taught her that she did not last at jobs very long, except for her present one, and so she was content to stay on even though the psychological test predicted poor ability and her production records verified the prediction. An opposite example is offered by the man who made an excellent score in a battery of tests for salesmen. He possessed all the traits and abilities considered desirable but had worked out very poorly on the job. An interview revealed that he was married to a wealthy woman and did not need money; he had taken the job in order to acquire a cloak of respectability. The fact that he did not sell as much as the next fellow was of little concern to him.

The practical expediency of costs often determines the extent of the test battery. For the average job, the cost of testing would be excessive if the test period required more than twenty or thirty minutes. Since all aspects of ability cannot be measured in this time period, the effectiveness of the test battery is necessarily lowered. Psychological tests are useful but limited in scope. We repeat this over and over, not to defend or attack testing, but to give a clear and honest understanding of the possibilities. Testing is desirable, but care must be exercised that testing, when applied in industry, is appropriate.

For example, some years ago an employer in a certain factory was interested in introducing a testing program and consulted one of this text's authors. The em-

ployer complained of his high turnover, long training period, and many other difficult problems. He wanted to introduce a battery of psychological tests so that his selection of employees would be improved. He now believed that testing would solve all his problems. A subsequent review of the type of work done in the factory indicated that psychological tests might be of value. In the discussions it was found that his labor supply was rather limited. A high-school education, either general or academic, was considered a basic requirement for hiring. The question was raised as to why a high school diploma was necessary, and why graduates of a general rather than a vocational high school were considered desirable. Probing revealed that the policy of the firm was to avoid union organization. In line with this policy, the factory never hired experienced people because they might be union members, and graduates of vocational high schools were not hired because it was believed that they were directly or indirectly connected with the unions. The point of view in this text is neither for nor against unions, and it is not the intention to introduce extraneous material during a presentation of psychological testing. However, this illustration serves to indicate that this employer wanted to use psychological tests to be able to continue avoiding unionization and still improve the selection of the applicants. This was called to his attention and he was advised to reconsider his so-called basic qualifications, since there was an available supply of experienced people. But he was adamant; he wanted the battery of tests to select the inexperienced people he considered hiring.

In many situations a person does not know what to do either because the problem he faces is too difficult or because there may be more than one correct solution. Similar situations arise in connection with hiring practices, and in such cases an employer may grasp at the idea of psychological testing as a drowning man supposedly grasps at a straw. If two applicants are available for a job paying $30,000 a year and both appear to be exceptionally good, the executive officer may prefer to "pass the buck" so that he cannot be blamed if the one employed does not prove to be as successful as predicted. At such a time he will believe that psychological testing or even fortunetelling might be good; for once the decision has been made, it is impossible to know whether the other applicant would have been any better. Hence a psychologist may be called in to give some tests and recommend hiring one of the two applicants. The extent of help a psychologist can give in such a predicament is limited. At best, he can only analyze and interpret the test results of the two candidates and point out any differences the tests reveal. He may not know—and in most instances cannot guess—whether the tests are related to successful job performance.

SOME DANGERS IN TESTING

The use of tests in the competitive and profit-conscious world of modern industry results in certain problems. One should always be aware of and on guard for these problems.

OVERPROMOTION A great danger in psychological testing is the possibility of its being oversold to a prospective client. Many psychological testing organizations exist, and some make overly enthusiastic claims to potential clients. When these organizations employ professional psychologists, the likelihood of exaggerated claims is not too great; but when they employ people who are immature professionals or pseudopsychologists, there may be trouble. Psychological testing is a good thing,

but like all good things it may be overdone. When unscrupulous individuals who claim to be psychologists fail to maintain rigidly high standards imposed by professional practice, then the use of testing can be harmful. All claims must be based upon substantial data, and they must be examined. The profession would rather have psychological testing proceed slowly than have it suffer a series of setbacks because of exaggerated claims or invalid uses. Precisely because of the generalized acceptance of tests, considerable caution must be exercised.

COMPETENCE OF THE TESTERS Before a business concern hires a psychologist or a professional organization to introduce a testing program, it should take the precaution of determining professional competence. As suggested in Chapter 1, the attainment of graduate degrees in psychology, membership in professional societies, and relative extent of experience should serve as guideposts. Costs should be related to the professional level of the staff and time spent in performing services. Further, the competent psychologist cannot guarantee absolute or positive results. All that he can do is indicate the possibilities and degree of expected success.

"SECRET" TECHNIQUES AND TESTS In the profession of psychology, and specifically in the use of psychological tests, secret formulas and secret techniques are highly unlikely. People who claim to use them are either unprofessional or unethical and should be so considered until they offer evidence to prove their claim. This evidence should be independently evaluated by professionals.

USING TESTS WITHOUT AMPLE REASON Previous sections have stressed the need for the statistical validity of a test to be established prior to its use. Sometimes all that is necessary for a test to become part of a selection procedure is for it to have a catchy name or to "look like" it should measure whatever skill is wanted (i.e., have face validity to the user).

Some years ago in New York City, finger dexterity testing was the vogue. One of the department stores used such a test to select packers, and within a short time many other stores followed suit. These other stores had no data but assumed that the first one knew what it was doing. Actually, this was not the case. The first store based its approval of the testing program on scanty research and an ambiguous report. Finger dexterity tests can be used with some success for certain purposes, but they have no value in the selection of packers. Blum and Candee (1941a) obtained correlations of about zero between the test results and the production records of a group of department store packers and wrappers. In a follow-up study the same authors found that speed and accuracy in clerical detail bore a closer relationship to employee production than finger dexterity (Blum and Candee, 1941b).

It is necessary to have an understanding of the background of psychological testing before using such tests. Otherwise one is not in a position to know the difference between a psychological test and a mere series of questions.

A joint committee of the American Psychological Association, the American Educational Research Association, and the National Council on Measurements Used in Education has published a set of recommendations in connection with test standards (American Psychological Association, 1954). Were all test developers and publishers to adhere rigidly to these recommendations, the consumer of tests would be on safer ground. The serious student of testing should be completely familiar with this set of standards.

ADVANTAGES OF TESTS IN INDUSTRY

The major advantage of psychological tests when used in industry is that the test is an objective and standardized behavior sample which lends itself well to statistical evaluation. Generally, it is easier to determine the value of a test than it is to evaluate other selection devices such as interviews, letters of recommendation, etc. Also, tests tend to be less subject to bias—particularly tests of aptitude and achievement. Tests of interest and personality (sometimes referred to as "inventories" rather than tests) may be more susceptible to either unconscious or deliberate bias on the part of the applicant under certain conditions. (This problem will be discussed in more detail in a later section.)

In most cases the cost of tests is not particularly high relative to the costs of other personnel evaluation aids, especially if the tests are of the paper-and-pencil rather than the performance type. A great deal of information about a person can usually be gathered in a relatively short period of time by using tests. Testing in industry is a useful aid in modern business and deserves serious consideration, despite the shortcomings mentioned earlier.

The great advantage in testing is that it can improve the selection process. The problem of hiring inexperienced people is faced by all firms to some degree. Such employees need many months of training. The cost of hiring plus the cost of training often total between $300 and $400 per employee, and in many cases may even total several thousands of dollars. If an employer expects to continue in a competitive business, such costs must be kept to a minimum. Psychological testing can reduce the costs of hiring people who will be successful, for its goal is measuring aptitude and predicting ultimate success among inexperienced applicants. However, as Brown and Ghiselli (1952) point out, it is not safe to assume that tests which best predict trainability will predict job proficiency equally well.

In an exhaustive review of all studies reporting the correlation between test performance and the two criteria of trainability and job proficiency, Brown and Ghiselli found only a slight tendency for a test that proves useful in predicting trainability also to be useful in predicting job proficiency, and vice versa. This means that factors important in learning a job may differ considerably from those that are important in maintaining proficiency on the job. After all, a bright person with certain skills may learn a job very quickly. However, if the job is not personally challenging or satisfying, he may terminate or perform in a manner considered unsatisfactory by his superiors.

Psychological tests can also be useful in the selection of experienced applicants. In many instances these tests serve as the only good check on the possibly exaggerated claims of an applicant. Some applicants successfully bluff their way through an application blank and an interview, but faking a test can be more difficult. Many girls who claim to have experience as power sewing machine operators base their claim on nonindustrial experience; there is a sewing machine in their home, and their mother has taught them to use it. They also claim to know the operations of the machine as well as various types of stitching. An interview cannot easily check such claims, but a standardized test can establish the accuracy as well as the speed of performance. Further, when norms are available the individual's ability can be judged more accurately than is possible when judgment is based solely upon an interview.

Psychological testing is also useful in selecting people with promotional ability, discovering reasons for job failure based upon personality traits, and even determining susceptibility to accidents, as we shall see in later sections of the text.

hedge on the various clauses in its contract that deal with the closed shop or general hiring practices. Many union executives believe that psychological testing serves only the interests of the employer; the fact that it advances the interests of the employee is open to debate. The unions' general disapproval of psychological testing programs must be reckoned with in a unionized industry.

Unfortunately, psychological testing has been subjected to more than its share of public criticism in recent years. It seems at times to be a favorite target for any group or person who needs a focal point to engender dissatisfaction While the use of psychological tests in the public school system has borne the brunt of these attacks, industrial testing has received its share. Examples of such critics are Hoffman's *The Tyranny of Testing* (1962), Black's *They Shall Not Pass* (1963), Gross's *The Brain Watchers* (1962), and Whyte's *The Organization Man* (1956). The last two authors were particularly critical of using tests to determine the fate of an individual as a job applicant. Whyte has even gone so far as to provide "tips" on how to get the "kind of score" psychologists "want" with their tests.

There is no question that tests are being misused and abused in some instances —indeed, we have already given examples and will give more later. However, test users today tend to be a rather competent lot, and employers themselves are becoming more sophisticated. Most test publishing houses maintain certain standards concerning who may purchase their tests. While these standards are sometimes difficult to enforce, they do discourage the incompetents from getting involved in testing. The major difficulty generally occurs with those pseudoprofessionals who have developed their own test which they peddle to anyone interested in buying it as a panacea. The real tragedy is that there are enough gullible management people to make the charlatans' efforts prosper.

As an example of this gullibility, one of the authors received a visit one day from a plant manager and his personnel director. Both were from a city several hundred miles distant. The reason for their visit was that a friend of the personnel manager who resided in the author's city had sent him a 15-question test of introversion-extroversion which he had cut out of the daily paper. Although the personnel manager had misplaced it before he could show it to the plant manager, the latter was so interested in its potential that both men had travelled several hundred miles in an attempt to get more information about the test. They wanted to use the test to screen applicants for their industrial sales force. Appalling as this illustration is of the degree to which management is all too often naive about tests and their appropriate use, it is only one of many such incidents that the authors have experienced.

Stagner (1958) has experimentally demonstrated the degree to which personnel managers can be "duped" by fallacious tests. He gave a group of personnel managers a published personnel inventory, but instead of giving each man his obtained scores, a fake "personality profile" was returned to each man with thirteen personality traits circled in red as being descriptive of that manager. As shown in Table 4.2, all managers had the same thirteen traits circled. These items had been collected from horoscopes, dream books, etc., and the thirteen circled items were chosen randomly.

Each manager was then asked to read over the items marked for him and decide how accurate each one was, using a scale from (a) amazingly accurate to (e) almost entirely wrong. The results showed that half of the managers felt their profile was an amazingly accurate overall description of them, 40 percent thought it was a "rather good" description, and the remaining 10 percent said their profiles were just average descriptions of their personality. The results for each item are given

TABLE 4.2 *Evaluations of Items by 68 Personnel Managers When Presented as a "Personality Analysis"*

Item	*Judgment as to Accuracy of Item Percent[a] Choosing*				
	a[b]	*b*	*c*	*d*	*e*
1. You have a great need for other people to like and admire you	39	46	13	1	1
4. You have a tendency to be critical of yourself.	46	36	15	3	0
5. You have a great deal of unused capacity which you have not turned to your advantage.	37	36	18	1	4
7. While you have some personality weaknesses, you are generally able to compensate for them.	34	55	9	0	0
9. Your sexual adjustment has presented problems for you.	15	16	16	33	19
10. Disciplined and self-controlled outside, you tend to be worrisome and insecure inside.	40	21	22	10	4
12. At times you have serious doubts as to whether you have made the right decision or done the right thing.	27	31	19	18	4
15. You prefer a certain amount of change and variety and become dissatisfied when hemmed in by restrictions and limitations.	63	28	7	1	1
16. You pride yourself as an independent thinker and do not accept others' statements without satisfactory proof.	49	31	12	4	4
18. You have found it unwise to be too frank in revealing yourself to others.	31	37	22	6	4
20. At times you are extroverted, affable, sociable, while at other times you are introverted, wary, reserved.	43	25	18	9	5
21. Some of your aspirations tend to be pretty unrealistic.	12	16	22	43	7
23. Security is one of your major goals in life.	40	31	15	9	5

[a] Not all percentages add to 100% because of omissions by an occasional subject.

[b] Definitions of scale steps as follows: *a.* amazingly accurate, *b.* rather good, *c.* about half and half, *d.* more wrong than right, *e.* almost entirely wrong.

SOURCE: R. Stagner. The gullibility of personnel managers. *Personnel Psychology,* 1958, *11,* 347–352.

in Table 4.2. The moral of the story, of course, is that you can impress anyone with the wisdom of your ability to assess personality if you keep your statements on a generally favorable level—a fact fortune tellers have been exploiting for hundreds of years.

ETHICAL STANDARDS OF TESTING

To prevent, as much as possible, the misuse of psychological tests, the ethical code officially adopted by the American Psychological Association (1959) devotes considerable space to test distribution and use. Some of the major aspects of this code are as follows:

1. The sale and distribution of tests should be restricted to qualified users. Who

is qualified will vary, of course, with the type of test and the background requirements of the psychologist.

2. Test scores should only be released to those individuals qualified to interpret them.

3. A person should not be subjected to testing under false pretenses. For example, the psychologist in industry may be placed in a conflict situation between loyalty to the person being tested and loyalty to his employer. In many cases he must guarantee anonymity to the person taking the test in order to get his cooperation, only to find pressure put on him from management to release the scores of certain individuals who were tested It is the responsibility of the psychologist to keep the testee clearly informed at all times of the purposes of the testing and the uses to which his scores will be put.

4. Tests should not be released for public use without adequate objective data to support their merit. If a test is being given for research (testing the test), this should be clearly indicated to all concerned.

5. Test manuals containing normative data and validity and reliability information should be made available for all tests. This information should be kept current.

6. Tests or parts of tests should *never* be published or presented in a public media such as newspapers, magazines, etc.

Besides the above, several additional standards might be stated.

7. All testing sessions should be carried out under standardized conditions with a qualified test examiner present.

8. One should never "prep" a person for a particular exam. To do so invalidates the test results, as the score received is no longer a true representation of the individual in comparison with those individuals who were not "helped."

INTRODUCING A TESTING PROGRAM

Chapter 2 discussed in some detail the basic steps necessary in the proper establishment of a selection program. Since a testing program is one type of selection program (remember that there are other kinds of predictors of job success), these steps bear repeating:

1. Examine the job in question.
2. Select a criterion.
   Select a predictor.
3. Measure job performance.
   Measure test performance.
4. Relate test performance to job performance.
5. Accept or reject the test as being of value.

The introduction of a testing program in a business organization demands that a research program accompany it. Attempts to select applicants with the aid of psychological tests based upon opinion and judgment rather than statistical facts are rarely, if ever, justified.

Before introducing a test battery, it is extremely advisable for employers to consult an industrial psychologist. He not only has training and background in tests but also has had experience with different testing programs. He knows that there is

no shortcut in the process, and that a testing program must be accompanied by research. There is a practical reason for research to parallel a testing program. A decision must be made as to which applicants will be recommended for hiring and which will be turned down. It is not at all safe to assume that the applicant with the highest score will be the best person for the job. For example, it has been unequivocally determined that in many occupations the people with the highest scores on an intelligence test are not necessarily the most successful employees; in many jobs scores above a certain point are conducive to inefficiency and high turnover. In many occupations the best employee, from the point of view of a prediction based upon an intelligence test score, is the one who ranks in the middle range. It is often necessary to establish a maximum test score above which it is unsafe to hire; a minimum score must also be established, below which it is unsafe to hire. In other words, test results must be compared with success on the job in such a manner as to determine adequately the statistical validity of that test as a selection instrument.

Another reason for employers to utilize the services of a psychologist is that some tests have misleading names: They can be used for selecting workers in occupations other than that implied by the name of the test For example, the Minnesota Clerical Test is a name-checking and number-checking test, but it has been found useful in the selection of inspectors and also department store packers. In both these instances the usefulness of this test was greater than when it was used to select clerical workers.

It cannot be expected that a battery of psychological tests which has been used successfully in one company will give the same degree of success in another. The type of personnel in two neighboring plants manufacturing the same type of product may vary tremendously. If known to the applicants, such factors as working conditions, morale, and selection technique used may predispose the better applicants to seek jobs in one plant rather than the other. Moreover, age, sex, skin color, and religion as a basis of selection are often artificial factors and may vitiate the validity of the testing program.

Because of relatively recent school experience, young people are often more familiar with pencil-and-paper tests and as a result do better on them. Older people may be frightened at the very thought of writing, as a result not only of passive decay but also of certain unpleasant memories connected with school tests. If factory A hires young people and gives a pencil-and-paper test, such a test may work for that factory. Factory B may be just as successful from the point of view of production but may hire older people who are equally efficient. This factory will probably not be able to use the pencil-and-paper test which was successful in factory A. Thus the principle that test batteries should not be borrowed is important. Naturally, this does not mean that one should not attempt to benefit from the experience of others, for such information may provide useful leads.

Also, as mentioned in Chapter 3, one may obtain differential prediction with the same test for different subgroups of people on the same job (moderator effects). In a study on the selection of sewing machine operators, Blum (1943) devised a zigzag pattern which the subjects were required to follow, using an actual machine with a needle but no thread. They were also required to sew a zigzag pattern in between two lines.

Before this job sample could be claimed to be a valid test for selection, it was necessary to test the test. In this particular study it was found that this sewing task

was moderately helpful in selecting experienced operators, but was practically useless in selecting inexperienced operators, that is, the test was a valid predictor of success for experienced applicants but not for inexperienced ones.

PROBLEMS OF ADMINISTERING TESTS IN INDUSTRY

A very important reason why a psychologist experienced in psychological testing should be employed for a test program in industry lies in the nature of the problems that arise in giving tests. Applicants or employees who take tests are likely to be cautious. They need encouragement in order to give their most typical performance, and they must be given honorable guarantees. Regardless of their success on the job, employees worry about the test, for they feel that poor test performance may cost them their job. If the precaution of dispelling such ideas is not taken, they are likely to show antagonism and resentment, and the employer will have a serious morale problem on his hands. Much of the anxiety of employees can be relieved by statements to the effect that the testing program to be used in the company in the future will depend upon whether the established employees do well on the tests, and that if the employees do not do well it will mean that the tests are unsuitable for the company. It is also desirable, when possible, not to give the employer the exact test results in any individual case; knowledge that individual test results will not be revealed gives the employees further reassurance.

The attitude of the examiner while administering the tests is important. He must be friendly and show a personal interest in the subjects but still maintain the standardized conditions for the test. He must know how to cope with unanticipated responses during the testing—laughing fits, talking, and even cursing. Ability to handle these so that they cause a minimum of interference is a prerequisite of an experienced examiner. The person administering the test program not only should be trained in psychological testing techniques but also should spend an apprentice period with an experienced examiner. This is true regardless of whether he has a master's degree in psychology or not. Some years ago one author hired an outstanding graduate student to help administer an industrial testing program. The student phoned him that the tests could not be given because of a broken stopwatch. The author's hurried taxi ride with a spare stopwatch proved to be unnecessary, for the student's watch merely needed winding. In his many courses the young man had never been taught to anticipate an unwound stopwatch.

CURRENT TESTING PRACTICES

The development and use of tests in industry are widespread at the present time, and they promise to continue on an increasingly broad scale. The remainder of this chapter will deal with just how extensive industrial testing is and will give examples of several testing programs and specific tests used in industry.

EXTENSIVENESS OF TESTING

To list all the companies in the United States that use psychological tests would be a lengthy and impossible task—impossible because published test results of studies are often described without mention of the specific companies involved. Moreover,

the companies that pay for research sometimes regard the material as "top secret" or as their private property, and so their policy is to discourage reports in the literature. Such a policy often gives a company an advantage over its competitors, and from this point of view the policy is wise. But at the same time it constitutes an obstacle in the development and use of tests An incomplete list of the companies which have used psychological testing can be made, however, and in some respects may be said to comprise an industrial *Who's Who.* This list has not been organized on the basis of either the type of company or the type of test problem because it was believed that a random order would best illustrate the wide and varied use of the tests. A random list of some of the companies and the tests used is as follows:

The Johnson & Johnson Company, well-known manufacturers of surgical bandages, have used tests to select foremen and clerical workers. The Vega Airplane Company and the Curtiss Wright Corporation use tests in their selection process. R. G. Le Tourneau, Inc., has developed a battery of tests to measure the primary mental abilities needed for job success with them. The Sun Oil Company has been interested in developing a battery which would be suitable for selecting people who could be trained to operate machine tools. The Lockheed Aircraft Corporation has had successful experience with personality, intelligence, and mechanical aptitude tests in the selection of many different types of employees. The Hawaiian Sugar Refining Corporation has found the use of tests for the promotion of employees successful. Aptitude tests as well as intelligence tests have been used by the A. C. Spark Plug Division of General Motors, the Woodward Governor Company, and Martin and Schwartz, Inc. The Harwood Manufacturing Company and the Kaiser Glove Company are two firms in the needle trades which have hired operators with the aid of tests. The American Aluminum Company selected a group of prospective foremen for training on the basis of psychological test scores. Apprentice toolmakers are hired by the Scovill Manufacturing Company with the aid of tests. The Milwaukee Electric Railway and Light Company uses a battery of many tests to select a category of employees known as "electrical trouble men." The Philadelphia Electric Company has found tests helpful in the selection of substation operators. Westinghouse Electric Company uses tests as part of its promotion program for clerical workers. The Atlantic Refining Company selects men for responsible positions, such as accountants and statisticians, with the aid of tests. Most of the insurance companies —among them Liberty National Life Insurance Company, Aetna Life Insurance Company, and Metropolitan Life Insurance Company—have found test batteries helpful in selecting agents.

The Harris Trust and Savings Bank of Chicago uses both personality and intelligence tests in its selection process. Considerable research has been done by the Kimberly-Clark Corporation; this concern has developed its own tests, such as the Kimberly-Clark packing and inspection test, which have been made available to the profession.

W. T. Grant hires clerical workers on the basis of tests. The Tremco Manufacturing Company has done much research on the selection of salesmen. R. H. Macy has used psychological tests for years in hiring employees for many of its departments. Procter & Gamble has found tests useful in selecting salesmen and sales managers. Aptitude tests have proved valuable to the Todd Company in hiring salesmen.

It is very difficult to obtain any really accurate set of figures indicating, in percentage terms, the frequency with which tests are employed in industry. One major

difficulty is that many firms may use tests only occasionally for specific openings which occur at infrequent intervals. Other firms may rely on some state or government agency such as the State Employment Service to do most of their testing and preliminary screening for them. Thus, one must really differentiate between formal and informal testing programs before attempting to determine the frequency with which industrial firms are using tests as a method of screening job applicants.

One such survey was carried out in 1959 by the *Industrial Relations News*, a trade magazine (1959). They surveyed 200 companies to see what percent used formal testing programs for job applicants. Sixty-five percent reported having such a program. Of those with a testing program, *all* agreed that their programs were highly useful and that they fully intended to continue them. When the nontest users were queried, 60 percent felt that testing could be useful, 2 percent felt testing was of little value, and 38 percent had no opinion. Thus, of the total sample only about 14 percent were not of the opinion that testing was of considerable value in an industrial setting.

SOURCES OF INFORMATION ABOUT TESTS

A major difficulty often encountered by a person wanting to install a testing program is where to find information about various kinds of tests—particularly information related to test use in industrial settings. There are a number of excellent sources for such information, and anyone who is apt to become involved in a testing program should be aware of these references.

Probably the best single source of test information is a series of *Mental Measurement Yearbooks* edited by Oscar K. Buros. Published periodically, Buros' books represent a continuing attempt to keep reference material related to tests as up to date as possible. They are a very complete source because they embody a *critical* survey of tests. They are not limited to industrial tests but include practically all work which has been done in the entire testing field. In each volume a group of contributors carefully selected by Buros reviews all the available material on practically all the psychological tests. For example, Buros' *Sixth Mental Measurements Yearbook* (1965) presents information on 1219 different tests and includes 795 original critical reviews.

Several extremely useful general textbooks are available which cover the general topic of psychological testing. These include Anastasi's *Psychological Testing* (1961), Cronbach's *Essentials of Psychological Testing* (1949), and Guion's *Personnel Testing* (1965). While they are not devoted exclusively to tests in industry, they present information on many of the tests currently being used in industry. In addition, they provide a great deal of general information on testing which is quite useful to a person involved in any sort of testing process.

A third source of readily available test information is provided by test publishers. While there are a large number of such sources, many of whom are quite small and who publish only one or two tests, the larger test publishers have catalogues which often are a convenient starting place for someone interested in establishing a test program.

The psychology journal *Personnel Psychology* has a section devoted exclusively to the exchange of information concerning industrial applications of various tests. It is called the *Validity Information Exchange* (VIE). Its purpose is to stimulate the publication of validity information in industry and to stimulate the use of such

information in guiding personnel actions. Every validity study is organized in the following succinct format for easy evaluation by the reader.

1. *Firm:* The sponsor for the study and the organization or firm within which the study was done (if not proprietary information).
2. *Problem and setting:* The problem to which the study was addressed is described, and the social, economic, organizational, and related elements in the setting for the study may be outlined
3. *Job title and code:* A job title and job code are given The title and code describe the job performed by the people who were studied. The code is taken from the United States Employment Service's *Dictionary of Occupational Titles* (Third Edition), Volumes I and II, U.S. Government Printing Office, 1965, or from supplements and new editions as they may become available
4. *Job description:* A description supplementing the job description given in the *Dictionary of Occupational Titles* is given when necessary
5. *Criteria:* The means by which job performance was observed, recorded, evaluated, and quantified are described.
6. *Sample:* The kind of people studied are described, such as by sex, age, educational level, job service, and marital status.
7. *Number of cases:* The number of people in the sample, the number of people in subsamples, and the uses of subsamples in the study are described.
8. *Predictors:* The kinds of data being investigated for their usefulness in guiding personnel actions are described. When tests are used as predictors they may be described by title and a reference number, such as "Miller Analogies Test (6·472)." This example refers to a description of the Miller Analogies Test found in review number 472 in Buros' *Sixth Mental Measurements Yearbook* which was described previously in this chapter Other references may be made to the fifth, fourth, and third yearbooks, to subsequent issues of the yearbooks as they may appear, or to bibliographies such as Buros' *Tests in Print* by the Gryphon Press, Highland Park, N.J., 1961
9. *Technique and results:* The methods used in evaluating the data are described and the results are summarized in statistical form.

Jones (1950) reviewed over 2100 references on employee selection and selected 427 studies as representing the "cream of the crop." The ten classifications of workers most frequently studied were: salesmen (75), clerical workers (60), teachers (49), assemblers (23), executives (23), inspectors (23), supervisors (21), typists (17), stenographers (14), and machinists (9). However, the important point of the article is that Jones found only eight reported studies which met the criteria of adequacy in both experimental design and report. So that the more serious student may refer to these studies, they are here listed:

Bellows, R. M. Studies of clerical workers, chap. 8. In W. H. Stead, C. L. Shartle, *et al.* *Occupational counseling techniques.* American Book, New York, 1940, 144–146 (study of coding clerks).

Blum, M , and B. Candee. The selection of department store packers and wrappers with the aid of certain psychological tests. *Journal of Applied Psychology,* 1941, 25, 76–85.

Guilford, J. P., and A. L. Comrey. Prediction and proficiency of administrative personnel from personal history data. *Educational and Psychological Measurement,* 1948, 8, 281–296.

Holliday, F. The relation between psychological test scores and subsequent proficiency of apprentices in the engineering industry. *Occupational Psychology,* 1943, 17, 168–185.

Otis, J. L., O. L. Endler, and L. E. Kolbe. Data-analysis methods, chap. 7. In W. H. Stead, C. L. Shartle, *et al.* *Occupational counseling techniques.* American Book, New York, 1940, 113–136 (study of department store salespersons)

Rundquist, E. A., and R. H. Bittner. Using ratings to validate personnel instruments: A study in method. *Personnel Psychology*, 1948, *1*, 163–183.
Sartain, A. Q. Relation between scores on certain standard tests and supervisory success in an aircraft factory. *Journal of Applied Psychology*, 1946, *30*, 328–332.
Selover, R. B. The development and validation of a battery of tests for the selection of clerical workers. *American Psychologist*, 1948, *3*, 291–292 (abstract) and personal communication.

Jones concluded by listing five requirements of a research report on employee selection using psychological tests:

1. Detailed job description, with each group treated separately
2. Complete description of the sample: *N* (sufficiently large); what proportion of the total population this represents and how selected, factors involved in hiring such as age, length of time on the job (preferably with widely differing employees treated as separate groups), and total experience in jobs of similar nature, and use of two samples, with one an applicant group.
3. Exact test titles; period in the employment experience when the tests were administered; whether the tests were a factor in hiring; where the tests were given; under what conditions and incentives the tests were given, reliabilities of tests with comparable groups.
4. Detailed description of the criterion; length of time on the job when the criterion measure was applied (with widely differing employees treated as separate groups); reliability of the criterion; some discussion of the validity of the criterion selected; if ratings were used, some estimate of the amount of contact the rater had with the employee; if production records were used, the duration of the period and whether there were any unusual factors operating at that time.
5. Adequate statistical treatment, with assurance that the assumptions governing the use of the given measures were met, and an actual report of the numerical results together with an appropriate measure of significance.

Dorcus and Jones (1950) have published a volume of abstracts which is essentially the record of the 427 "cream of the crop" studies referred to in Jones's study. It includes over 200 occupations listed alphabetically from accountant to YMCA secretary, and reports such items as subjects, tests used, criterion, validity of results of the study, and reliability in the criterion. A sample abstract follows:

Surgent, L. V.: The use of aptitude tests in the selection of radio tube mounters. *Psychological Monographs*, 1947, *61* (283), p. 40.

1. Subjects: 233 radio tube mounters, female. Tests used as part of application procedure but results not used in selection.
2. Tests: (1) Minnesota Rate of Manipulation Test: Placing, (2) Minnesota Rate of Manipulation Test: Turning, (3) O'Connor Finger Dexterity Test, (4) Purdue Pegboard, (5) O'Connor Tweezer Dexterity test.
3. Criterion: a single over-all rating based on the pooled judgment of the supervisor of training and 2 instructors; 5-point scale; rating during training period.
4. Validity: biserial correlation between test scores and criterion (groups divided into satisfactory and unsatisfactory):

Test	*Correlation*	*t*
1	.56	6.56
2	.50	5.79
3	.48	5.60
4	.64	7.39
5	.59	6.81

Multiple correlation of test scores and criterion (optimum order of test addition):

Test	*Correlation*
4	.64
5	.72
1	.76
3	.76
2	.76

Follow-up of 35 of the 233 mounters· prediction of criterion scores on basis of a regression equation correlated with over-all rating by immediate supervisor on the job·
Two-test equation, involving Tests 4 and 5 .60
Three-test equation, involving Tests 4, 5, and 1· .43

5. Reliability: rank-order intercorrelations ($N = 51$) of 2 over-all fitness rankings of quality and quantity of work produced in training (latter 2 one week after first):

Items	*Correlation*
Quality and quantity	.66
Quality and over-all fitness rating	.74
Quantity and over-all fitness rating	.85

Note the similarity between the format used by Dorcus and Jones in their *Handbook* and that currently used by the VIE.

In summary, the preceding sources of test information provide an excellent introduction and information base relating to test data, standardization, and research.

SPECIFIC EXAMPLES OF TESTS

In the following pages examples are given of different types of tests used today in industry. No attempt at a complete listing can be made because of limited space. Rather, the intent is to provide the reader with a feel for the typical test, using as examples some of the more commonly used test instruments of each type.

MENTAL ABILITY TESTS

There are a number of intelligence tests which have either been adapted or specially developed for use in industry. The primary function of most of these tests is to serve as a preliminary screening or evaluation device without taking too much time. Most such tests are designed to be completed in 15 or 20 minutes, and the items included provide as much face validity as possible. Some of the more frequently used tests will be briefly discussed.

OTIS TESTS OF MENTAL ABILITY Probably the most widely used industrial screening tests are the *Otis Self-Administering Tests of Mental Ability* (1922–1929). These

consist of two series of four equivalent tests. The first series is designed for high school and beginning college individuals, while the second series is designed for grades four to nine. In industry these tests have been used for selecting applicants for almost every conceivable type job. The success of the tests, as one might suspect, varies greatly from situation to situation (Dorcus and Jones, 1950). They generally are more useful with lower-level jobs, since they do not discriminate well at upper ability levels (Anastasi, 1961).

WONDERLIC PERSONNEL TEST This test is an abridged adaptation of the Otis tests (higher series). It has a time limit of only 12 minutes, making it particularly attractive to industrial users. In spite of its shortened form, it correlates above 0.80 with the original Otis test. The term *personnel* has been used in the title to reduce the threatening characteristic of any test that appears to deal with intelligence. Since the items on the Wonderlic are those which were found to differentiate between good and poor workers on a variety of industrial jobs, it is not surprising that the test often is found to be a better predictor than the Otis, in spite of its shortness. Excellent normative data are presented in the Wonderlic manual. As with the Otis, its validity varies greatly depending upon the situation, but its highest validities have occurred in the selection of workers for clerical jobs. Parallel-form reliabilities of 0.82 to 0.94 have been reported (Wonderlic and Hovland, 1939).

THURSTONE TEST OF MENTAL ALERTNESS One of the interesting features of this test is that it provides two subscores, verbal (V) and quantitative (Q), plus a total score (Thurstone and Thurstone, 1943). The test has a time limit of 20 minutes, and consists of 126 verbal and quantitative items in alternate order and ascending difficulty. Reliabilities near 0.90 have been reported. It is available in two forms and appears to have its greatest validity in sales and clerical jobs.

THE ADAPTABILITY TEST One of the few selection instruments which was developed specifically for industrial research, this test has a 15-minute time limit and also has been named to avoid a negative reaction by applicants to having their mental ability measured (Tiffin and Lawshe, 1942). It provides good face validity. It has been used primarily with clerical workers and first-line supervisors.

MECHANICAL APTITUDE TESTS

Tests dealing with mechanical aptitude can be classified into two subgroups: measures of mechanical reasoning and/or information, and measures of spatial relations. The reason the latter tests are included is that many mechanical tasks involve the rapid manipulation of pieces and parts relative to one another. Thus the ability to perceive geometric relationships between physical objects and to be dexterous in the manipulation of such objects is generally perceived as being part of the general complex of mechanical aptitude.

MINNESOTA SPATIAL RELATIONS TEST This was designed to test both dexterity and spatial relations (Paterson, 1930). A picture of the test is given in Figure 4.1. As the picture indicates, the test consists of a series of 58 geometric shapes which

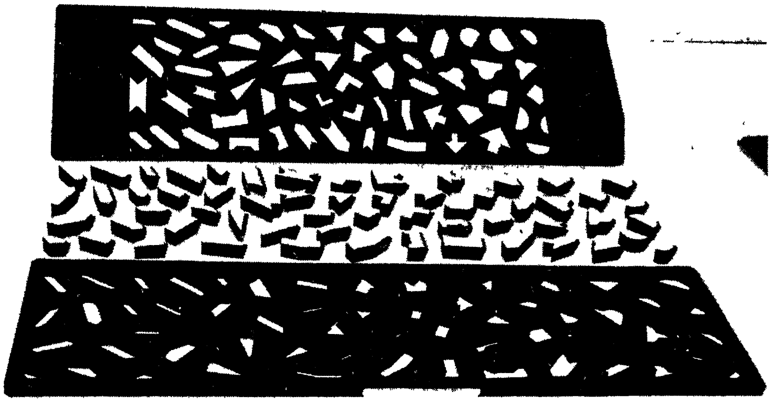


Figure 4.1. Minnesota spatial relations test (Courtesy Educational Test Bureau Division, American Guidance Service, Inc., 420 Washington Ave S. E., Minneapolis, Minn.)

have been cut from a large board. The task is to place the cutouts into their appropriate slots as fast as possible. Both amount of time and number of errors are scored. Four boards are used (A, B, C, and D) with the same set of blocks fitting boards A and B and another set fitting boards C and D. The first board serves as a warm-up for the following three boards which are used to obtain the person's score.

MINNESOTA PAPER FORM BOARD TEST This test is a paper-and-pencil instrument, thus eliminating dexterity as a variable in performance (Likert and Quasha, 1941–1948). It consists of 64 multiple-choice items. Each item presents the parts of a geometrical figure "cut up" into pieces, followed by five assembled geometric forms. The task is to select the form which represents what would be obtained if the cutup parts were assembled. Reliabilities for this test are in the mid-eighties. Ghiselli (1942) obtained a correlation of 0.58 between test scores and the success of inspector-packers. The test does *not* appear to be highly correlated with either intelligence tests or with other mechanical aptitude tests.

MINNESOTA MECHANICAL ASSEMBLY TEST This test is a revision of an earlier test (Paterson, 1930). It is designed as a test of mechanical knowledge and reasoning, and also involves dexterity and spatial relations ability. The test is of the work-sample type, wherein a person is given a series of disassembled mechanical contrivances (e.g. a clothespin, a doorbell, etc.) and asked to reassemble them. The total test comes in three boxes, with 11 objects in each box. The entire test takes nearly one hour. Reliabilities range from 0.72 to 0.94.

BENNETT TEST OF MECHANICAL COMPREHENSION This is one of the most widely used of all tests of mechanical knowledge (Bennett, 1940). A paper-and-pencil test, it uses pictures to present the testee with various questions about mechanical facts and principles. Separate forms are available for trade school applicants, engineering applicants, and women. Reported reliabilities range from 0.77 to 0.84.

TESTS OF SENSORY ABILITY

Success in certain types of jobs is often found to be related to such physical attributes as visual acuity, color vision, and sensitivity of hearing. Several standardized tests or test instruments are currently available for evaluating the degree to which a person possesses each of the above abilities. Such tests are commonly referred to as tests of sensory ability.

TESTS OF VISION Vision is a rather complex ability consisting of many different types of visual skills. Thus we can distinguish between the ability to see things far away (far acuity) and the ability to see things close to us (near acuity). People also differ in terms of how well they can perceive relative distances (depth perception) and how well they can differentiate colors (color vision). Any or all of these abilities may be found to be important for job success.

Probably the best-known test of far visual acuity is the *Snellen Chart,* a display consisting of rows of letters which gradually diminish in size from top to bottom of the display. The person is usually placed at a distance of 20 feet from the chart and asked to read successive lines until he is no longer able to distinguish the letters. A ratio is then computed which expresses his acuity as follows:

$$\text{Acuity} = \frac{\text{distance at which person } i \text{ can read the line of print (usually 20 feet)}}{\text{distance at which the average person can identify the same line}}$$

Therefore, if the smallest line which person *i* can read at 20 feet is a line which most people can read clearly at 80 feet, his score is 20/80. Normal vision is, of course, 20/20.

The Snellen Chart is generally considered to be only an approximate measure of a person's total visual ability. Several general purpose acuity instruments are available commercially which measure a variety of visual skills. The most widely used instruments of this type are the Ortho-Rater (Bausch and Lomb), the Sight-Screener (American Optical Company), and the Telebinocular (Keystone View Company). All three instruments are general purpose in that each provides measures of depth perception, near and far acuity, phorias (a phoria exists when the eyes fail to converge naturally to an appropriate focal point distant from the individual) and color discrimination.

Probably the best-known test of color blindness is the *Ishihari* test. It consists of a series of cards or plates which have designs comprised of colored dots. The designs are constructed so that the background dots consist of one color while the design is made up of another color. The colors were deliberately selected so that a person with normal color vision will perceive the correct figure (usually a number),

while the color-blind person will perceive either no figure or an incorrect design. Appropriate responses to each plate for normal vision, red blindness, green blindness, and total color blindness are given in the test manual.

TESTS OF HEARING Auditory acuity, like visual acuity, has many dimensions. However, the most common dimension is simply the degree to which a person is sensitive to sounds of different pitch. The usual instrument employed to measure auditory sensitivity is the *audiometer*. This generates pure tones of different frequencies which can be varied in terms of their loudness by the examiner. The individual being tested listens for the tone, which is gradually raised in intensity. When he hears the note he signals to the examiner who records the intensity. Usually on half the trials the examiner starts with a loud tone and decreases the intensity until the testee signals he can no longer hear it, while on the other half he starts below the audible point. The average of a series of such trials is taken as the person's *threshold* for that frequency. Figure 4.2 shows the general relationship that exists between the

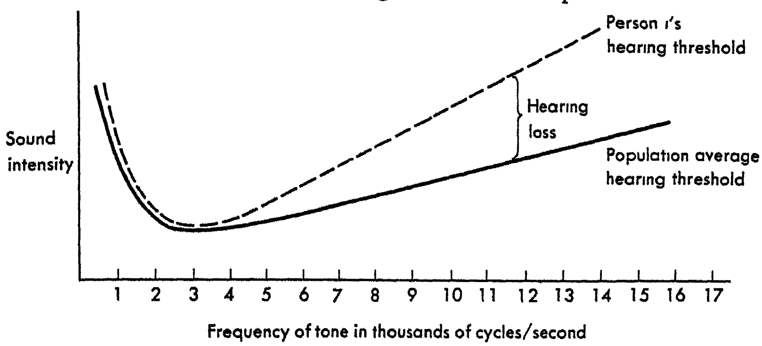


Figure 4.2. General functional relationship existing between the pitch of a tone and the auditory threshold (amount of sound intensity necessary for that tone to be audible) for the population average An acuity curve for a person with a hearing loss at the upper frequency levels is shown for comparison.

frequency of a tone and the threshold for that tone. Notice that we are most sensitive to sounds in the two-to-three-thousand cycles/second range. For comparison, a hearing curve for a person having a substantial hearing loss in the upper frequency levels is also included in Figure 4.2. The amount of loss at any particular frequency is the distance between the population average and the individual's own curve.

Using the audiometer it is possible to prepare a chart, known as an *audiogram*, which presents a clear profile of a person's auditory acuity. One such audiogram or profile is prepared for each ear.

TESTS OF MOTOR ABILITY

Quite a variety of standardized performance tests are available for measuring fine and/or gross motor coordination. Some of the frequently used tests are listed below.

PURDUE PEGBOARD This timed test requires pins to be placed into small holes in the pegboard, using first the right hand, then the left hand, and finally both hands together (Tiffin, 1941). In the second part of the test, the pins must again be

placed in the holes but this time collars and washers must be assembled with tl pins as they are placed (both hands are required for this part of the test). The tw parts of the test are designed to measure manual dexterity and fine finger dexteri respectively.

CRAWFORD SMALL PARTS DEXTERITY TEST There are two parts to this timed te (Crawford and Crawford, 1946). In part 1, pins are placed in holes with tweezer and then metal collars are placed over the pins. In part 2, a screwdriver used to screw down small screws after they have been placed by hand into thread holes.

MINNESOTA RATE OF MANIPULATION TEST Like the Crawford Test, this test consis of two discrete parts. In the first part the task is to place 58 round blocks into board having an equal number of holes. The blocks are 1½ inches in diamet and are slightly thicker than the holes are deep, i.e., they still protrude after b ing correctly placed The first part is referred to as the "placing" test. The secor part consists of having the subject turn all the blocks over and replace them their holes. This is called the "turning" test. A person's score is the time requir to complete each task. An interesting aspect is that the two parts do not correla very highly. Correlation between part scores is generally in the order of 0.50.

INTEREST INVENTORIES

The likelihood of success in an occupation in which one is interested should l greater than the likelihood of success in an occupation which a person finds n particularly interesting. For this reason it would seem sensible to obtain son measure of a person's interests prior to placing him on a job or suggesting a particular vocation. Of course, the most obvious way to ascertain a person's i terest is simply to ask him what he is interested in. However, for a variety of reason the direct approach has proven to lack the necessary reliability. All too often peop lack sufficient knowledge about specific vocational choices to make even an i telligent estimate as to whether or not they would really find it an interesti career. Instead, they can merely express their interest in terms of gross estimat such as "That sounds like it might be fun, from what little I know about it."

For this reason, interest inventories have been developed over the years to a complish the assessment in a more subtle and sophisticated manner. The develo ment of most such inventories has followed the same basic pattern with only min variations on the overall general theme. The procedure for developing such te might be as follows:

1. Prepare a variety of items dealing with how well one likes all kinds of act ities, ideas, objects, and types of people. Some typical items might be:

   I like to go swimming.
   I enjoy going to concerts.
   People who talk a lot are fun to be with.
   Writing school reports is interesting work.

2. The items are then given to people in different occupations or occupation groups, and people in these occupations are asked to indicate whether these ite describe *them*, that is, do they agree or disagree with the statement.

3. *Keys* are then developed for each occupation using those items which statistically discriminate that occupation from all others, that is, people in that occupation respond to it significantly more (or less) often.
4. These keys can then be used to get an interest score for each occupation for a person taking the test. To the extent that a person checks those items on a particular key, he is considered to have interests similar to people already in that occupation. The implicit assumption is that the more similar the interest pattern, the more apt one is to succeed in that occupation.

STRONG VOCATIONAL INTEREST BLANK (SVIB) Developed by E. K. Strong, Jr., the SVIB is made up of 400 items divided into eight different parts (1938). There are 47 different occupational keys available for use with the men's form of the test and 28 keys for scoring the women's form. Most of the keys are for higher-level occupations. (Relatively little success has been achieved in deriving keys for low-level jobs.) In addition to the keys for the various occupations, the SVIB also provides a number of scoring keys which can be used to evaluate interest maturity, occupational level, masculinity, and specialization level. The latter scale is relevant for today's highly complex job world, since it contains items which have been found to differentiate between the interests of men who have specialized in a given field and those who have remained "general practitioners." The SVIB scales have reliabilities in the mid-eighties.

KUDER PREFERENCE RECORD A somewhat newer interest inventory is the Kuder Preference Record. Its two most frequently used forms are the Vocational Form (Kuder, 1934) and the Occupational Form (Kuder, 1956). The former is made up of a number of forced-choice alternatives in which the person must select from three listed activities the one he would like most and the one he would like the least. The scales on the Vocational Form include outdoor, computational, mechanical, persuasive, scientific, artistic, literary, musical, social service, and clerical. The Occupational Form was constructed in a manner similar to the SVIB and has 38 different occupational keys.

PERSONALITY INVENTORIES

More controversy has surrounded the use of personality measures as predictors of job success than with any other type of test. As a rule, the validity of such instruments has tended to be somewhat less than impressive, although at times rather substantial predictability has been obtained. Numerous reasons have been suggested to account for the general lack of success in industrial situations of personality inventories, such as lack of reliability, inappropriateness of context, and susceptibility to faking. Even those tests developed specifically to counter these objections have failed to produce validities of the magnitude one might wish. The authors are of the opinion that "personality" contributes to a significant degree in determining the job success of many individuals, particularly in higher-level jobs. Nevertheless, successful personality assessment for a specific job is the exception rather than the rule.

Measures of personality may be conveniently classified into either self-reporting questionnaires or projective tests. Self-report inventories generally involve responses to a question in terms such as how much the testee "agrees" with the question, how

well it "describes" him, or which of *k* alternatives he "prefers most." The essence of the response is that of self-description. Typical test items are:

Do you worry over possible misfortunes?
Does discipline make you discontented?
At a reception or tea do you feel reluctant to meet the most important person present?
Do you often worry about your health?

Projective tests, on the other hand, usually involve obtaining an unstructured or "free response" from the individual. For example, the test may provide an incomplete sentence and asking the testee to finish it in any way he wishes. Another major difference between self-description inventories and projective measure is that the scoring and interpretation of the latter are much more complex and require much more training. Examples of both types of personality measures are described in the following sections.

GUILFORD-ZIMMERMAN TEMPERAMENT SURVEY This test is a combination of three earlier personality inventories (Guilford and Zimmerman, 1949). It was developed using a statistical procedure known as factor analysis, a technique which groups things (in this case test items) into homogenous clusters which are usually relatively independent of one another.[2] The Guilford-Zimmerman gives a score on ten different factors. Each trait or personality factor is based upon thirty different items, and each trait is supposedly independent of the other traits. The traits are as follows:

G.	General Activity	O.	Objectivity
R.	Restraint	F.	Friendliness
A.	Ascendance	T.	Thoughtfulness
S.	Sociability	P.	Personal Relations
E.	Emotional Stability	M.	Masculinity

The reliability of the different scales range from 0.75 to 0.85.

MINNESOTA MULTIPHASIC PERSONALITY INVENTORY (MMPI) Probably the most widely known self-description inventory is the Minnesota Multiphasic Personality Inventory (MMPI). It consists of 550 affirmative statements which the person taking the test must classify as *true*, *false*, or *cannot say* (Hathaway and McKinley, 1943). Using standard item analysis procedures in which a control group of normals had to respond significantly more (or less) frequently to an item than a corresponding group of clinical patients in order for that item to be considered discriminating, ten different "clinical scales" have been developed as follows:

Hs.	Hypochondriosis	Pa.	Paranoia
D.	Depression	Pt.	Psychesthenia
Hy.	Hysteria	Sc.	Schizophrenia
Pd.	Psychopathic deviate	Ha.	Hypomania
Mf.	Masculinity-femininity	Si.	Social introversion

[2] For example, suppose one had an intelligence test partly made up of vocabulary items and partly of arithmetic items. A factor analysis of the relationship between these items would probably indicate two distinct factors or clusters. This would indicate that the verbal items have a lot in common, as do the arithmetic items, but the verbal items have little in common with the arithmetic items.

In addition to these scales, the MMPI also provides four "validity" scales which can be used to indicate the extent to which the respondent understood what was required of him and cooperated in taking the test.

The test has had limited industrial use, particularly because of its length and possibly because of clinical terminology. However, it has been used in governmental agencies and has been the subject of congressional inquiries related to "invasion of privacy." The reliabilities of the subscales tend to be quite low—thus, one may be investing a great deal of time for some data of uncertain consistency.

BERNREUTER PERSONALITY INVENTORY Like the Guilford-Zimmerman, the Bernreuter Personality Inventory is a combination of previous personality measures (four, in this case). It consists of 125 yes-no items and gives scores on the following scales:

B1N.	Neuroticism	B4D.	Dominance
B2S.	Self-sufficiency	F1C.	Confidence
B3I.	Introversion	F2S.	Sociability

The first four scales were developed by Bernreuter. However, they were found to be highly correlated (e.g., B1N and B3I correlated as high as 0.95) and therefore somewhat redundant. Flanagan (1935) subsequently factor-analyzed the test items and found two relatively independent scales, F1C and F2S. These latter keys should be used in place of, rather than in conjunction with, the first four scales.

RORSCHACH TEST This test is a projective instrument rather than an inventory (Rorschach, 1947). The Rorschach test presents a series of ten standardized inkblots to the subject. In each instance the subject states freely what he sees, either in the parts of each blot or in the complete blot. The examiner, who must be trained in this technique, analyzes the results in numerous ways. Scoring is based on the type of response—whether the subject reports movement, human figures, animate or inanimate objects, etc. The rigid supporters of this technique exhibit an almost religious fervor in their claims, and conversely the attackers insist that the test is useless. The authors believe that the test does have diagnostic value in analyzing personality and emotional aspects of the individual, but it does depend upon the skill of the test interpreter.

THEMATIC APPERCEPTION TEST (TAT) A second well-known projective test, the Thematic Apperception Test (TAT) was developed by Murray (1943). A series of pictures is presented to the subject and he is asked to tell an extemporaneous story about each picture. Analysis of the themes of the stories, especially when the same themes are repeated, reveals dominant drives as well as conflicts and inhibitions. In this test as well as the Rorschach, special training is necessary.

HUMM-WADSWORTH TEMPERAMENT SCALE A personality inventory, the Humm-Wadsworth Temperament Scale consists of 318 questions to which the testee responds *yes* or *no* (Humm and Wadsworth, 1935). Seven different aspects of temperament are obtained from the inventory. Very little validity information has come to light regarding its value in industry.

ACTIVITY VECTOR ANALYSIS A highly controversial test is the Activity Vector Analysis. According to its author, W. V. Clarke (1953), it is used in over 100 companies in business and industry. The test consists of a sheet containing 81 descriptive words. The respondent "places an X (in column 1) before every word that has ever been

used by anyone in describing you." In column 2 the respondent "places an X opposite every word which you honestly believe is descriptive of you." This results in a scoring of four vectors—aggressiveness, sociability, amiability, and avoidance—and an overall score or activity level. The novelty is the claim that after only ten days of training any person who is a college graduate and has had five years of industrial experience can become an analyst. We wonder whether the popularity of this test is based upon "You too can become an expert and in only ten days."

A STUDY IN CONTROVERSY

Perhaps no single test better illustrates the controversial problems involved in industrial testing than does the Activity Vector Analysis. A brief review of the controversy surrounding this assessment device should serve to illustrate some of these difficulties. Locke and Hulin (1962) did a comprehensive review of the use of this test as an industrial selection device. They reviewed a total of 18 studies, 17 of which were relevant to the value of AVA in industry. The majority of these studies, however, were of the concurrent validity type in which the AVA was given to present employees. Locke and Hulin reported that there was evidence that it could differentiate between

1. Managers and production workers
2. Executives and a mixed worker group
3. Seven different occupational groups ranging from machine operators to company presidents
4. Good and poor employees in several different occupations

However, in all the present-employee studies, validities were obtained by developing a key based upon that sample and then applying the key back to the sample upon which the key was developed—a procedure known as "fold-back" validation.[8] Rarely was any attempt made to conduct the necessary process of cross-validation on a different group of present employees. Locke and Hulin believe that all too often the authors of the AVA studies exceeded accepted limits in discussing the value of their findings, often leaving the reader with the impression that their validities were indicative of the *predictive* power of the test.

Only one study, in the opinion of Locke and Hulin, used a true predictive validity procedure where a previously developed scoring key was given prior to employment and then checked later for validity. In this instance no validity was obtained.

It would appear that a decision that the AVA has predictive value must at least be postponed until more substantial evidence is available. At issue here, however, is not only the actual value of the test, but also the discrepancy between the claims made for its value and the actual empirical findings. Locke and Hulin are not alone in taking the AVA to task on the basis of misrepresentation of research findings. Dunnette and Kirchner (1962) have strongly protested the "colored" reporting of AVA data by Meranda and Clarke (1959). Bennett, in his review of the AVA for Buros' *The Fifth Mental Measurement Yearbook* (1959) is also rather uncom-

[8] The term "fold back validation" in general refers to the procedure of validating an empirically deprived scoring key using the same data from which the key was obtained. This is clearly not a good procedure and one which will lead to artificially high validities.

plimentary, stating, "The mumbo-jumbo of allegedly sophisticated statistical procedures is no substitute for demonstrated validity." The interested reader may enjoy the Dunnette and Kirchner article (1962) and the corresponding reply by Meranda and Clarke in the same *Journal of Applied Psychology* issue.

OVERVIEW OF PERSONALITY TESTING IN INDUSTRY

Ghiselli and Barthol (1953) reviewed 113 studies dealing with the validity of personality inventories in employee selection. Their results are summarized in Table 4.3. They conclude that under certain circumstances scores on personality inventories correlate better with proficiency on a wider variety of jobs than might have been expected. The authors recognize both the potential value of personality testing in industry and the need for a vast increase and improvement in research and development.

TABLE 4.3 *Weighted Mean Validity Coefficients of Personality Inventories for Various Occupational Groups*

Mean r	*Total Number of Cases*	*Total Number of r's*	*Occupation*
0.14	518	8	General supervisors
0.18	6433	44	Foremen
0.25	1069	22	Clerks
0.36	1120	8	Sales clerks
0.36	927	12	Salesmen
0.24	536	5	Protective workers
0.16	385	6	Service workers
0.29	511	8	Trade and crafts workers

SOURCE: E. E. Ghiselli and R. P. Barthol. The validity of personality inventories in the selection of employees. *Journal of Applied Psychology*, 1953, *37*, 18–20.

EXAMPLES OF INDUSTRIAL TESTING

A few of the many studies that have been conducted will be reported rather completely to illustrate the systematic discipline required in the use of tests for industrial purposes. None of these studies is to be regarded as illustrating an ideal testing situation; each was conducted within the limitations imposed by the particular industrial situation, but each attempted, to the degree possible, to conform to the rigid standards of research.

BLUM STUDY

A study on the use of dexterity tests to select watch-factory workers was conducted by Blum (1940). The first step in the research called for job analysis of the different types of work performed in the factory. This brought to light the follow-

ing important job requirements: fine finger movements, the manipulation of tweezers, and the ability to continue performing delicate and sometimes intricate tasks over long periods of time without any increase in tension or maladjustment.

A review of the available test material indicated that in view of the factory's requirement of a brief testing period, the Johnson O'Connor finger dexterity and tweezer dexterity test would probably be best. Prior to embarking on the testing program, however, the various available criteria of success were surveyed and three measures were obtained: length of employment, salary ratio, and foremen's ratings.

The length of employment criterion was divided into four categories. The first was the "less than one week" group. The majority of employees in this category left or were dismissed within one week because of poor ability for the job. From the point of view of this criterion, this group was the poorest. The second category was the "one week to four months" group. Management believed that the job could be learned adequately within four months. Those who could not learn it in this time either were dismissed or resigned. From management's viewpoint, such an employee was inefficient; according to the employee, he could not earn enough. In any event, the employment relation was severed. The third category was the "four months to one year" group. These employees were regarded by management as moderately successful. The fourth and final category was the "one year or longer" group. These employees were considered to be most successful; the training costs for them were lowest and the group had the highest production.

The second criterion was salary ratio. Since all employees were on a piece-rate basis, earnings were a direct reflection of production. The figure used was the average of the weekly earnings over a three-month period. This time was considered long enough to average out minor peaks or slacks in business and minor difficulties in production. The actual criterion was expressed in the form of an index, with X dollar earnings per week equal to a base of 100.

The third criterion was the rating of employees by the foremen. Each foreman was asked to give an overall rating on the basis of his opinion of the employee's usefulness and efficiency. In accordance with this rating, each employee was classified as excellent, good, average, poor, or unsatisfactory.

Five measures of test performance were established. Two were based on speed (total time required to do each test). Another measure was the rating of the subject's qualitative performance on each test. Since this adds meaning to any objective score such as total time or items completed, it will be described in detail in the hope of encouraging the use of such a measuring technique.

The qualitative rating is the examiner's overall estimate of the test performance. It includes how well the subject follows directions, tension during the test, and the method used to complete the performance. A rating scale of "good," "average," and "poor" was used for the following qualities: accuracy of selection, grasp of pins, positioning of pins, placing of pins, hand tremor, conditions of board, pace, position and movements of arm, and body posture.

The fact that there is an appreciable distribution in qualitative performance ratings is illustrated by the findings in this particular study as presented in Table 4.4.

This measure of test performance results in an overall rating by the examiner of the subject's manner during the test. Most performance tests and many pencil-and-paper tests readily lend themselves to such a rating. These ratings are useful in considering a person for hiring.

The fifth measure was a comparison of the time taken to do the second half of

Personality and interest tests (inventories) are intended to measure personality characteristics or patterns of interests of individuals, on the assumption that such characteristics or interests may be related to performance on various kinds of jobs. When such tests are used as employment tests they are typically used in the prediction of job performance, in essentially the same manner as aptitude tests. It should be added here that there are certain serious limitations to the usefulness for employment purposes of presently available personality and interest tests. These limitations will be discussed in a later chapter.

TESTING THE TESTS—THE EXPERIMENTAL APPROACH

It is of utmost importance that any tests that are used, for employment purposes or otherwise, be "validated." In general, the concept of validity refers to the extent to which a test is capable of achieving the aims or purposes it is intended to serve. Actually, there are different types of validity; these will be discussed later in this chapter. It is only when a test has been demonstrated to have an acceptable degree of validity that it can be used safely, with reasonable assurance that it will serve its intended purpose. The process of test validation is sometimes referred to as "testing the tests."

There are two general methods of validating the tests that are commonly used in personnel testing. One of these, the "present-employee" method, consists of testing present employees and correlating the test results with an appropriate criterion of job success. The other, the "follow-up" method, consists of testing new employees at the time of hiring, filing the test results, and later determining the relationship between the test results obtained at the time of hiring and the criterion measures of the employees after they have been on the job for a period of time.

Each of these methods has certain advantages and disadvantages. A long-range testing program, however, preferably should make use of both methods if it is to yield maximum benefit.

In addition to these methods of validating tests, other procedures are occasionally used. Brief mention will be made of certain of these methods later on.

The present-employee and follow-up methods involve certain steps, some of which are alike. A discussion of these steps will clarify the similarities and differences.

Present-Employee Method of Testing the Test This method of testing the test (test validation) involves the use of a group of present employees on the job or jobs in question.

1. Select battery of experimental tests. An early step in a test-validation project is that of selecting a battery of tests to be tried out. These

tests should be chosen in terms of the extent to which they are considered to measure attributes that are judged to be important to job success. This selection preferably should be made on the basis of information obtained from a job analysis.

2. *Administer tests to present employees.* The tests selected are then given to the present employees on the job. This is usually done on a voluntary basis, in order not to create too much unrest among the employees.

When a plant begins, for the first time, a project of test validation that involves testing a large group of present employees, the question of the reaction of the employees naturally arises. Will they become anxious or unduly excited? Is there a possibility of the testing program causing worker unrest? In this book will be summarized research work and test results that have been obtained by testing hundreds of employees in numerous plants. Many of these plants were strongly unionized. In the union plants, several labor organizations have been represented. In no case has any trouble arisen either with employees or union representatives. Wherever the management has signed a union contract, representatives of the union are contacted and the project explained to them before actual testing is begun. Usually a mimeographed slip reading somewhat as follows is given to each employee:

> The Personnel Department is conducting a series of experiments. You have our assurance that this testing is being done to "test the tests" and that the results will *not* be used, now or later, in any way that will affect your standing with the company.
>
> PERSONNEL DIRECTOR

It should be emphasized that in using this method of testing the test, management should be utterly sincere in its plan to use the results *only* to test the tests. Under no circumstances should the test results be used as a basis for transfer or layoff, or for any other purpose detrimental to the employee or which he may *believe* to be detrimental. After a testing program has been thoroughly tried out and installed, many uses for the test scores may be made, uses that will be advantageous to both employees and management; but in the "testing the tests" stage of the program, management should adhere scrupulously to its agreement not to use the results for any other purpose than the one stated.

3. *Select appropriate criteria.* At some early phase of a test-validation project it is necessary to determine what criterion of job performance to use. The criterion was discussed in Chapter 2, so we need not repeat that discussion here. The point will be re-emphasized here, however, that the criterion or criteria used in any test-validation study should be *relevant,* meaning that is should reflect the standards by which the performance of employees should be evaluated in terms of management's objectives.

4. Obtain criterion information on present employees. After determining what criterion or criteria to use, it is then in order to obtain criterion information on the individual employees now on the job. Depending on the criterion in question, this may involve the accumulation of available records (such as production records, sales volume, and the like), or it may consist of obtaining ratings from supervisors on job performance of the employees, or other appropriate processes.

Depending on the manner in which the results are to be analyzed, the criterion information may be used to divide the total employee group into two groups such as "high" (or above average) and "low" (or below average), or it may be expressed in quantitative terms such as units of production or numerical ratings.

5. Analyze results. After the test scores and criterion information have been obtained, the results may be analyzed in several different ways. One method is to divide the total group of employees into a "high" criterion group and a "low" criterion group as mentioned above. Next, the employees are divided on the basis of test scores into two test-score groups. For each test-score group the per cent of "high" criterion employees is then computed. If statistically significant differences in these percentages are found for the different test-score groups, it may be safely assumed that the test is measuring something of importance on the job. The results of this type of analysis usually are shown in an "expectancy chart" such as illustrated in Fig. 5.1. This type of chart shows the "odds" of being successful of the applicants within each test-score group.

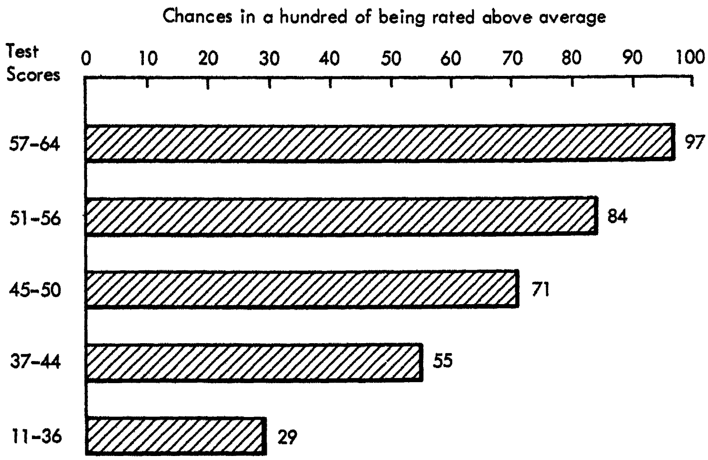


Fig. 5.1. Expectancy chart showing the relation between scores made on the Minnesota Paper Form Board (see page 162) and rated success of junior draftsmen in a steel company.

Another, and even more effective, method of determining the relationship between test results and job performance is to compute a coefficient of correlation between the test scores of the employees and their criterion values. This has certain advantages over the expectancy-chart method mentioned above. In the first place, it gives a more accurate indication of the *amount* of the relationship between test scores and efficiency. In the second place, it enables the employment manager more effectively to take advantage of the all-important selection ratio (see p. 132) in using the test. In the third place, it makes possible the computation of the relative importance of several tests in an employment battery so that the tests may be "weighted" according to their importance. Finally, the use of the correlational method makes it possible to offset, statistically, whatever influence such factors as experience on the job, or age, may have had on the test scores of the employees. Further mention will be made later of these possible influences.

Although the correlational approach has certain advantages over the expectancy-chart method, these are largely statistical advantages. In terms of understanding by nonstatisticians, the expectancy-chart method probably cannot be excelled.

Follow-up Method of Testing the Test As indicated earlier, this method consists of giving the test, at the time of employment, to a large number of employees whose placement has already been decided upon by previous, non-test methods of selection and placement. These employees usually are hired just as they have always been hired, but before they have been placed in their jobs, or, in fact, even before they have been told they have been selected for employment, they are required to take the battery of experimental tests. These tests at this time have no effect whatever upon whether the applicants will be hired. The steps involved in this method are described below.

1. Select battery of experimental tests. This step is essentially the same as with the present-employee method.

2. Administer tests to applicants. The tests are administered to applicants who are to be employed for the job in question, but the applicants should not know at the time that a decision has been made to employ them. The test results are then filed until a later date.

3. Select appropriate criterion. This determination is made in the same way as with the present-employee method.

4. Obtain criterion information on the new employees. The criterion information on the new employees usually should be obtained after sufficient time has elapsed for them to demonstrate their actual abilities on the job. Usually this would be after completion of training, or at least after the completion of most of the training.

5. Analyze results. This step is carried out in the same manner as with the present-employee method.

Comparison of the Two Methods of Testing the Test Each of these methods of test validation has certain advantages and disadvantages. For example, one possible objection to the present-employee method when validating aptitude tests is that a test may be measuring some ability that is improved significantly by experience on the job. In other words, it may be measuring more achievement than aptitude. In validating aptitude tests by this method it is therefore necessary to be sure that the tests, besides dividing the employees according to their ability on the job, do not show a significant correlation with experience on the job. In other words, it is necessary to be sure that the employees who score high on the test are not scoring high simply because they have had an opportunity to learn the skill being tested. Whether this is the case may be determined by correlating the test scores with experience on the job. If the test is to be used later for employment and/or placement it must satisfy two requirements. First, it must show a positive correlation with ability on the job. Second, the scores on it must not be appreciably related to experience on the job. When these two conditions are satisfied, it may be assumed that the test is *not* measuring something that is improved markedly by experience on the job, but that it *is* measuring something that is necessary for adequate performance on the job. This reasoning applies only to *aptitude* tests, not to *achievement* tests. When tests of the latter type are used, a significant correlation between test scores and experience is to be expected and in no way reduces the value of the tests.

If the test scores show *some* correlation with experience, the net relationship between test scores and job performance, after the effect of experience has been eliminated, can be determined by partial correlation. The procedure for computing partial correlations may be found in any standard textbook of statistics.[2]

Although the influence of experience on test scores in such a situation can be eliminated by the use of partial correlations, a more straightforward manner of getting around such a problem is through the use of the follow-up method. In this case the test scores are obtained *prior* to any experience on the job, and therefore cannot be influenced by job experience.

The present-employee method has certain other possible disadvantages. In some instances, the *present* employees on a job may represent a highly select group, since most of those who were *not* satisfactory either were separated or left of their own accord. In such a case, the correlation between test scores and criterion values would *not* represent the relationship that one would expect with the follow-up method.

Further, the "mental set" of present employees toward taking tests

---

[2] J. P. Guilford, *Fundamental Statistics in Psychology and Education* (3rd ed.) (New York: McGraw-Hill Book Company, 1956).

on a voluntary, "experimental" basis may be different from the "mental set" of applicants. This difference in set can influence performance on some types of tests, especially personality and interest tests. Where this influence is of some consequence, it would be preferable to validate the test by the follow-up method. If this is done, the test is then validated in the same type of situation as the one in which it will later be used.

Another disadvantage of the present-employee method is that the arrangements for testing present employees sometimes are difficult to work out, especially since it is necessary to take people away from their jobs in order to test them.

Where any of the above conditions exist or might have important influences on test results with the present-employee method of test validation, the follow-up method probably would be preferable. The follow-up method, however, requires more time. Frequently, it does not result in any proved tests until the program has been under way for several months or, in some cases, several years. Should the first battery of tests selected for tryout prove entirely unsatisfactory, it is necessary to start the whole procedure over again. This involves the loss of a great deal of time.

Fortunately, in practice the employment manager need not decide upon one or the other of the methods discussed and then limit himself exclusively to that particular method. He may proceed with both methods simultaneously. When this is done it is possible to produce a working battery of tests almost at once (by using the present-employee method of testing the tests) and to obtain still more evidence of the value of these tests as new employees are tested at the time of placement. This statement may seem to contradict itself because it may be argued that if a test is used for placement there will be no variation in the scores on this test among the new employees placed on any given job and, therefore, it will be impossible to correlate the scores with later success on the job. Actually, even though test scores are used as a basis of employment or placement, there will ordinarily be a sufficient spread of scores on the test, even among employees who are hired for a specific job, to make possible the subsequent statistical evaluation of the test in terms of the follow-up method. Though the variability of the group of employees placed on any given job will be smaller than that of all the applicants, nevertheless the variability of all the applicants and the variability of the applicants who are placed on the job will be known and it will therefore be possible to determine on the basis of the follow-up method just how much the test in question increases the efficiency of the placement procedure for the particular job.

The point to be emphasized throughout this discussion is that no one —whether he is an employment manager, a psychologist, or anyone else— can predict with certainty which tests will be desirable tests for placement on any particular job. Often a rather accurate estimate can be made; but

it is usually desirable to check this estimate against correlations between the actual performance of the employees on the job and the scores that they made on the tests at the time of placement.

Other Methods of Test Validation Although the present-employee and follow-up methods of test validation are by all odds the most commonly used procedures, and in general the most defensible methods, there are certain other procedures that can be used.

One of these we might call a "job-group" method. In this method the test scores of employees on each of two or several jobs are compared to ascertain whether there are significant differences. In one company the Bennett Test of Mechanical Comprehension [3] was administered to employees on each of several jobs. Following are the mean test scores for those on certain jobs.

Job	*Mean Test Score*
Insulators	30 9
Pipefitters	33 8
Electricians	36.4
Welders	39.7
Instrument mechanics	42 4

The use of this method is predicated on the assumption that by "natural selection" people in an organization tend to "gravitate" into the kinds of jobs that are reasonably in line with their abilities. If it can be assumed that, in general, most persons on a job are achieveing a reasonable degree of success on that job, it is then possible to *compare* the employees on various jobs in order to ascertain what *differences* there are from job to job, as in the above example. It is then possible to select for initial placement, or for promotion, those individuals whose test scores most nearly correspond with the test scores of individuals now on the job.

Another procedure that will be mentioned is that of "synthetic" validity. This term was originally used by Lawshe [4] ". . . to denote the inferring of validity in a specific situation. The concept is similar to that involved when the time study engineer establishes standard times for new operations, purely on an *a priori* basis through the use of 'synthetic times' for the various elements constituting the operation."

The procedures involved in synthetic validity studies are complex and time consuming, and will not be discussed in any detail. Suffice it to say that the procedures depend upon two essential points: (1) the identification, in reasonably objective terms, of work elements or other job characteristics—which is essentially a job analysis process; and (2) the "validation" of a test *on numerous jobs with a common work element* or *job characteristic.* If, in general, it is found that there is a relationship

[3] See Appendix D for distributor.

[4] C. H. Lawshe, "Employee Selection," *Personnel Psychology*, 5 (1952), 31–34

between *scores on a given test,* and *performance,* on *jobs with a characteristic in common,* it is then possible to use that test in the placement of other individuals for *other* jobs with that *same* characteristic. This procedure probably offers greatest potential usefulness where there are not enough present employees or applicants available for more conventional test-validation procedures.

In a study of clerical jobs, Lawshe and Steinberg [5] found that this procedure offers reasonable promise. They used the *Job Description Check List of Clerical Operations* [6] to identify the work operations of 262 positions in 12 companies. The people in these positions were given the *Purdue Clerical Adaptability Test,*[7] which results in several subscores. By a procedure that need not be discussed here, various operations had been determined to have "critical" requirements on each of the several subscores on this test, such as "spelling." In general, it was found that the larger the number of "critical" work operations in a job (such as on spelling), the higher were the scores of individuals on such jobs, as shown in Fig. 5.2. This implies that there is a tendency for employees to

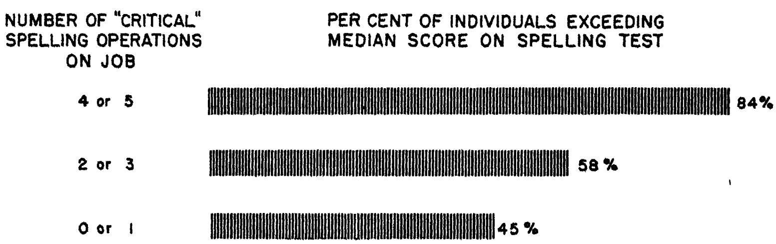


Fig. 5.2. Per cent of individuals exceeding median score on spelling subtest by number of "critical" spelling operations in their jobs. (Adapted from Lawshe and Steinberg, *op. cit.*)

"gravitate" to those kinds of jobs for which they have some particular aptitude. In the present illustration, it would be reasonable to place on jobs that have been *objectively* determined to have several operations with "critical" spelling requirements, those persons who have high spelling-test scores.

In another study by Cunningham and McCormick,[8] the Worker

[5] C. H. Lawshe and M. D. Steinberg, "Studies in Synthetic Validity. I. An Exploratory Investigation of Clerical Jobs," *Personnel Psychology,* 8 (1955), 291–301.

[6] Available from Southworth's Extension Service, West Lafayette, Indiana.

[7] Available for purchase from the University Book Store, 360 State Street, West Lafayette, Indiana.

[8] J. W. Cunningham and E. J. McCormick, Factor Analyses of "Worker-Oriented Job Variables," Occupational Research Center, Purdue University, Lafayette, Ind., June, 1964: Prepared under contract with the Office of Naval Research (Contract No. Nonr-1100(9)).

Activity Profile was used in the analysis of 400 jobs. (As indicated in Chapter 3, this is a job analysis format that provides for checking or rating 161 job activities and related characteristics.) By a procedure that need not be described here, it was possible, strictly on the basis of these analyses of the jobs, to derive "attribute indexes" for each of a number of human attributes, such as aptitudes and motor abilities. These indexes, in turn, were correlated with separately derived job requirements (for corresponding attributes) developed and reported by the U. S. Employment Service.[9] Some of these correlations are given below:

Attribute	*Correlation*
Aptitudes	
Verbal	.66
Numerical	62
Spatial	.43
Form perception	.43
Clerical perception	.57
Intelligence	.64
Motor abilities	
Motor coordination	.37
Finger dexterity	.48
Manual dexterity	.46
Eye-hand-foot coordination	.39

While these correlations are far from perfect, they are at least suggestive of the possibility of being able to establish job requirements in a fairly objective manner by the use of a reasonably objective job analysis procedure. Although the present status of the development of synthetic validity does not yet permit its widespread use, the general approach seems to be reasonably promising.

FACTORS DETERMINING THE FUNCTIONAL VALUE OF PERSONNEL TESTS

There are several factors of both a practical and theoretical nature that determine the functional utility of personnel tests in a business or industrial organization. These include the *validity* and *reliability* of the tests used, the *selection ratio,* and the *percentage of present employees who are "satisfactory" on the job.*

Validity In a broad sense, the "validity" of a test refers to the degree to which the test is capable of achieving the aims or purposes it is intended to serve. As indicated above, the process of "testing the test" is one of determining the validity of a test for the purposes of selection or placement. We have seen that the *degree* of validity, as reflected by a correlation or by an expectancy chart, affects the *extent* to which the test

[9] *Estimates of Worker Trait Requirements for 4000 Jobs,* U.S. Employment Service, (Washington, D.C.: U.S. Government Printing Office, undated).

can be expected to serve its intended purpose of selection or placement.

Although we are primarily interested in the validity of tests for selection and placement purposes, it may be useful for us to discuss briefly the various specific *types* of test validity that are appropriate in the use of tests in various situations, such as education, vocational counseling, and clinical practice, as well as in personnel work. There are basically four types of validity, these being related to different aims or purposes of testing.[10]

Content validity. Content validity is evaluated by showing how well the *content* of a test samples the subject matter or kinds of situations that it is intended that the test measure. It applies particularly to achievement tests, or tests given in schools. Generally speaking, this determination needs to be made on the basis of the *judgments* of "experts" with regard to the *appropriateness* of the questions or problems (the "content" of the test) as measures of that which the test is supposed to measure.

Predictive validity. Predictive validity is evaluated by showing how well *predictions* made by the test are *confirmed* at some *subsequent time.* The follow-up method of test validation is a typical illustration of this type of validity, since the validity of the test is determined at some time after the test has been given. Usually such validity would be established by determining the relationship between test scores and a criterion of performance, such as performance on a job.

Concurrent validity. Concurrent validity is evaluated by showing how much the scores on the test in question are related to some *present* (concurrent) criterion of performance or status. The present-employee method is an example of this type of validity, since test scores are correlated, or otherwise compared, with a performance criterion. The "job-group" test-validation method is also an example, but it involves the use of a "status" criterion rather than a performance criterion.

Construct validity. Construct validity is evaluated by determining what psychological *qualities* a test measures, such as "introversion," or "intelligence." Such validity frequently is determined by correlating one test with another that has been previously found to measure the quality in question, or by the use of a statistical process called "factor analysis" which can identify the extent to which various tests measure a human quality in common. The synthetic validity procedure is also concerned primarily with construct validity.

As indicated above, the present treatment of personnel tests in industry typically involves the concepts of *predictive validity* and *con-*

---

[10] A thorough discussion of the concepts of validity and reliability of tests is included in "Technical Recommendations for Psychological Tests and Diagnostic Techniques," *Supplement to the Psychological Bulletin,* March, 1954, Vol. 51, No. 2, Part 2, available from the American Psychological Association, 1333 Sixteenth Street, N.W., Washington 6, D.C. Price $1.

current validity. In connection with both of these concepts of validity, a criterion of job performance usually is involved, although job-status (job group) criteria sometimes can be used. Various criteria have been discussed previously on page 41, and it is important at this state to keep these criteria in mind. Unless one has a relevant criterion for the employees used in the validation study, he cannot even discuss the *predictive* or *concurrent* validity of a test.

Reliability The reliability of a test is the degree to which a test measures *consistently* whatever it does measure. Actually, there are different types of reliability; three have been differentiated by the American Psychological Association.[11] Measures of these are referred to as:

Coefficient of stability
Coefficient of equivalence
Coefficient of internal consistency

The methods of determining these three types of reliability are described below.

The *test-retest* method is used to derive a *cofficient of stability*. With this method, a group of subjects are given the same test twice, with a time interval between. The two sets of scores are then correlated. If memory or practice from the first administration can affect significantly the scores of individuals on the second administration, this method is not appropriate. To put it another way, this method is most suitable for use with tests on which memory or practice effects are of limited consequence.

The *alternate-forms* method (used to derive a *coefficient of equivalence*) is one in which two separate, but equivalent, forms of the test are given to the subjects. Such forms typically include items that are similar in nature, but different in actual content. The scores on these two forms are then correlated to determine the degree of reliability. If the test-retest method cannot be used because of memory or practice effects, the alternate-forms method may be suitable in case there are two alternate forms available.

The *split-halves* method is one procedure for deriving an *index of internal consistency*. By this method, a test is given to a group of subjects under standard administrative conditions. Each test is scored, however, on each of two halves of the items, and these two sets of scores are then correlated. The two "halves" may be chosen randomly ("chance" halves), or they may consist of alternately allocated items ("odds vs. evens"). This method is most suitable when there are numerous items in the test, and where the test is essentially one of "power" rather than "speed." (A speed test is one in which each item can be accomplished by most subjects, but where speed is important, as in a dexterity test. A power test is one in

---

[11] "Technical Recommendations for Psychological Tests and Diagnostic Techniques," *Supplement to the Psychological Bulletin*, 51 (No. 2, Part 2, March, 1954).

which the *ability* of the subject to complete the test items is more important than his *speed.*)

It should be added that there are other ways of determining or inferring test reliability, but the above methods are the ones most commonly used. It should also be pointed out that there is no such thing as *the* reliability of a test; rather, reliability is in part a function of the group of individuals tested. In the use of personnel tests, it is desirable to determine the reliability with the individuals on whom the tests are validated. This is not usually done, however. Rather, a person using a test will rely on reliability information based on previous groups of subjects, as reported by the test publisher.

Importance of reliability. One might ask this question: Of what importance is test reliability? This question might be answered by saying that if a test is highly reliable, it is possible to put greater reliance on the scores that individuals receive on the test, than if the test is not very reliable. Since, in most employment situations, one has only a single score on a given test for each applicant, it is useful to know how reliable the test is in order to know how stable such scores are. Generally speaking, the reliability of most tests is considered to be satisfactory if the coefficients of reliability are in the .80's or .90's. There are some circumstances, however, in which tests with lower reliabilities might still be used.

A test may have high reliability and yet be quite worthless for any given employment situation. A test might measure height, weight, or even general intelligence with high reliability and yet show little or no correlation between test scores obtained and success of the employees on the job. In other words, the test might have high *reliability* but low *validity* (predictive or concurrent validity), or even no validity at all. But if a test has a *low* reliability, it is not likely to have a satisfactory validity. This is simply another way of saying that no test is likely to have a higher correlation with anything than that test will have with itself. Thus a test that on repeated testing gives scores that correlate with the first testing only .45 is not likely, except by chance, to relate higher with production or any other criterion of employee desirability. It is primarily for this reason that the psychologist using tests in industry insists that they be reasonably reliable before he makes any attempt to determine their validity for particular situations.

Relationship Between Reliability and Validity A question that is often raised when the validity or reliability of tests is under discussion is: How reliable and valid must a test be in order to be worth while? This is both a reasonable and a natural question when one considers that our measurements of both reliability and validity are usually expressed in the form of a coefficient of correlation that makes possible a variation in the degree of reliability or validity all the way from zero (or even a minus quantity, if we wish to be theoretically exact) to 1.00. If we remember

that reliability in a test is necessary because it limits the validity, we may phrase the question more simply by asking: How high must the validity coefficient of a test be for the test to be worth while?

The answer to this question depends upon the use that is being made of the test. The user of tests is nearly always interested in one of two objectives, but is seldom interested in both at the same time. Either he is interested in making a careful and accurate aptitude analysis of *each person tested,* which is to be used for individual prediction or vocational guidance, or he is interested in selecting from a large group of individuals a smaller group that, *on the average,* will surpass the larger group in some particular respect. In individual counseling work, which deals with vocational aptitude and guidance, the psychologist is interested in the first-named objective. His work will stand or fall on the accuracy of his predictions for individual clients. He therefore has little use for aptitude tests that do not have a validity sufficiently high to justify their use in individual prediction. The exact value the validity of a test should have, to meet this requirement, is not completely agreed upon by all students of the subject. It is uniformly agreed, however, that the higher the validity of the test the better and that *there is no substitute for validity for individual prediction.*

On the other hand, one may be interested in segregating from a large group of persons tested a smaller group that, on the average will surpass the larger group in whatever trait is being tested. This is, in fact, the situation that confronts the employment manager. He is willing to accept, on the basis of tests, a few individuals who will fail on a given job, and to reject (or place upon some other job) a few who, if placed upon that job, would have succeeded, *if on the whole his percentage of successful placements is higher with the tests than it is without them.* In other words, although the employment manager would like to have every swing result in a home run, he usually recognizes that under normal circumstances this is not practical, and is therefore satisfied if he can improve his batting average. Under these circumstances the validity of the tests can be much lower. But one may still ask: How low can it be? A categorical answer to this question can be given, but the full significance of the answer will be clear only after a thorough study of the next section, which deals with the *selection ratio.* The answer is that a test probably will be valuable, no matter how low the coefficient of validity, if it indicates *some* relationship between test scores and the criterion; or, in statistical terms, if there are fewer than five chances in a hundred that the obtained coefficient of validity could have resulted from the operation of chance factors alone. Often this rule will admit tests whose validity is as low as .30 or even lower. The use of tests with such low validity is sufficiently contradictory to much current thought among psychologists to warrant a fairly detailed justification for the conclusions reached above.

The Selection Ratio Given a personnel test that has a validity coefficient indicating *some* relationship with the criterion, and given more employees to be placed than can be placed on the job in question, the functional value of the test to an employment manager depends upon the ratio of those placed to those tested who are available for placement. This has been referred to as the "selection ratio." [12] An example will clarify the operation of this principle.

If a certain test is given to a large number of employees for whom a criterion of successfulness as employees is available, and if the scattergram of test scores against the criterion is plotted, the points will ordinarily fall into an oval-shaped area somewhat similar to the oval in Fig. 5.3. The higher the coefficient of validity, the narrower will be the

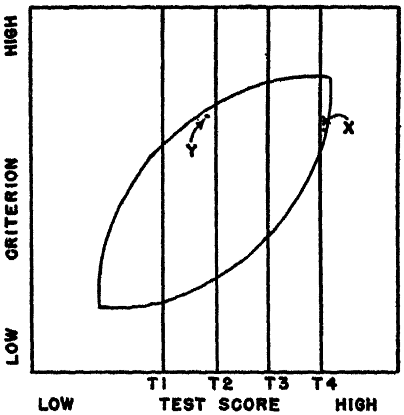


Fig. 5.3. Effect of shifting the critical score required of applicants on average criterion score of employees hired.

oval; and the lower the validity, the more nearly the oval will approach a circle. A validity coefficient of approximately .60 will result in a scattering of scores approximately covering the oval area shown in Fig. 5.3. Now, if employees are placed without regard to test scores, their criterion scores usually will be the average of all individuals falling within the oval. If only those are placed upon this job who have test scores as high as or higher than T_1, those not placed on the job will clearly have, on the average, lower criterion scores than the group as a whole, and those placed will accordingly be higher in their criterion scores, on the average, than the group as a whole. A still higher average criterion score for the

[12] H. C. Taylor and J. T. Russell, "The Relationship of Validity Coefficients to the Practical Effectiveness of Tests in Selection: Discussion and Tables," *Journal of Applied Psychology*, 23 (1939), 565–578.

group placed can be achieved by setting the critical test score at T_2. By moving the critical score to T_3, T_4, or even higher, still more favorable placements, according to average criterion score, can be made.

If a given number of persons, say 60, are to be placed, any one of the conditions mentioned above may exist; which one exists will depend upon the selection ratio that is utilized: that is, the ratio of the number placed to the number tested. Suppose we work with a ratio of 1.00—that is, all those tested are placed. In this case, the distribution of test scores will be over the whole range of possible test scores; the criterion scores will be over the whole range of possible criterion scores; and the test will contribute nothing whatever to the efficiency of the placement procedure. Now suppose that we test 80 individuals and place the 60 who score highest on the test, either not hiring the 20 who score lowest or placing them on some other job. We thus reduce those placed to 75 per cent of those tested, or reduce the selection ratio to .75. Under these conditions, we will place on this job only individuals who test at least as high as T_1, and the average criterion scores of those so placed will clearly be higher than the average of the group as a whole. By testing 120 persons and placing the 60 who score highest on the test, the selection ratio will be reduced to .50 and only individuals to the right of T_2 will be placed. The average criterion score of this group will not only be higher than that of the whole group, but will also be higher than that of the group placed when the critical test score was at T_1. Thus, by increasing the number tested before the 60 to be placed are identified, the selection ratio will be decreased with a continuous increase in the average criterion score for the group of 60 finally placed. If, for example, the plant is expanding so greatly that 600 new employees can be tested before 60 are selected for location upon this particular job (or if the labor market were such that 600 applicants were tested before 60 were hired for this job), the selection ratio would be decreased to .10, only those testing at least as high as T_4 would be placed upon the job, and the average criterion score of the group of 60 placed under these circumstances would be much higher than the score of 60 placed under any larger selection ratio.

The foregoing discussion is of course based on the assumption that the placement of employees is successful in proportion to the average success of the employees placed. Anyone can readily see that even working with a selection ratio of .1, some individuals (like X in Fig. 5.3) will be placed who will be poorer according to the criterion than a few other individuals (like Y in Fig. 5.3) who have not been allocated to this job. But if one is willing to measure the success of the testing program by average results rather than by individual cases, the results will be more and more favorable as the selection ratio is decreased.

We have already stated that psychologists dealing with vocational guidance and individual consultation are usually more interested in

making accurate individual predictions than in making group predictions. Most of these psychologists have tended, therefore, to evaluate a test almost entirely in terms of its validity coefficient. They have stressed the fact (which is unquestionably true for individual prediction) that there is no substitute for high validity; that if two tests have been validated against the same criterion, and one has a higher validity coefficient than the other, there is no way to make the one having the lower validity serve as well as the other. The main point of the discussion is that in group testing, where one is interested in average rather than individual results, one can make the test with the lower validity perform as well as the other *by sufficiently reducing the selection ratio.* In other words, in group testing, *a reduction of the selection ratio is a substitute for high validity.* This statement does not mean that this substitute will work if the test has no validity at all, but it does mean that if the test has any significant validity, however small, it is possible for the employer to get the same functional value from it that he could get from a test of any validity, however high, if he is able sufficiently to reduce the selection ratio.

A practical objection to the principle of increasing the efficiency of a test by decreasing the selection ratio may be raised—namely, that there is a limit to the number of applicants who can be tested before the desired number are placed. It is true that, for numerous reasons, an employment manager is seldom able to test 600 or even 200 men before 60 are placed. But it should be remembered that a ratio of 1:10 can be achieved by placing one person out of ten tested just as well as by placing 60 out of 600 tested. One does not need to wait for a great expansion in hiring before advantage can be taken of a reduced selection ratio.

The objection may also be raised that advantage cannot be taken of a reduced selection ratio unless there are more applicants than there are jobs to be filled. This is true if all persons employed are to be placed upon the same job. But almost always an expansion of plant personnel involves hiring for several jobs, not just for a single job. Therefore, the advantage of a reduced selection ratio usually can be achieved even when, as in a period of emergency production, there is difficulty in getting *enough* applicants to fill the jobs. Even when all applicants are hired, placements can be made on various jobs in such a way as to take advantage of individual differences by means of reduced selection ratios. *A reduction in the selection ratio can be utilized whenever two or more employees are being placed on two or more different jobs, if tests of some validity are available for each of the jobs.*

Percentage of Present Employees Considered Satisfactory A further factor that affects the efficiency of a personnel test in a given employment situation is the percentage of present employees who are considered satisfactory. This factor may be made clear by reference to Fig. 5.4. Suppose

available a test with a validity coefficient of .50 and are using a selection ratio of .50. Table 5.2 shows the increase, due to using the test, over the percentage of satisfactory employees prior to the use of the test. The values in Table 5.2 were obtained from the Taylor-Russell tables reproduced in Appendix B. If only 5 per cent of employees placed by traditional means are successful, then the expected increase to 9 per cent represents an 80 per cent increase in the number of satisfactory employees placed by the test, under the specified conditions of test validity and selection ratio. For larger percentages of satisfactory employees that have been achieved without the test, the percentage of increase achieved by using the test becomes increasingly smaller. If 90 per cent of employees placed by traditional means have been successful, the increase of this percentage to 97 per cent by the test, used under the specified conditions, results in an improvement of only 8 per cent in the number of employees satisfactorily placed.

Table 5.2

INCREASES IN PERCENTAGE OF SATISFACTORY EMPLOYEES PLACED ON A JOB OVER VARIOUS ORIGINAL PERCENTAGES OF SATISFACTORY EMPLOYEES WHEN A TEST WITH VALIDITY COEFFICIENT OF .50 IS USED WITH A SELECTION RATIO OF .50

A	B		
Percentage of Satisfactory Employees Placed on the Job without *the Test*	*Percentage of Satisfactory Employees Placed on the Job* with *the Test*	*Difference in Percentage Between Columns* A *and* B	*Percentage of Increase of Values in* (B) *over Values in* (A)
5	9	4	80
10	17	7	70
20	31	11	55
30	44	14	47
40	56	16	40
50	67	17	34
60	76	16	27
70	84	14	20
80	91	11	14
90	97	7	8

The general conclusion is that, other things being equal, the more difficult it has been to find and place satisfactory employees without using test procedures, the greater the gain one may expect from a suitable testing program.

Use of Taylor-Russell Tables The purpose of the foregoing discussion is to make clear that several factors, each relatively independent of the

others, operate to determine the functional value of a selection test to an employment manager. If the employment man knows these factors —that is, if he knows the validity of his tests, the selection ratio with which he is working, and the percentage of present employees considered satisfactory—he can predict definitely just how much he will improve the placement process by using the test. And if this amount of improvement is not satisfactory, he can improve his placement by almost any reasonable amount if the selection ratio can be decreased.

Fig. 5.5 reproduces a chart that shows how the percentage of employees selected who will be successful is determined by the validity of the test and the selection ratio. This chart deals with an employment situation in which 50 per cent of present employees are considered satisfactory. The base line in this figure gives the selection ratio, and each of the curves plotted indicates a different test validity. It will be seen that by using a test with a validity of .90, and by reducing the selection ratio to .60, the percentage of satisfactory employees placed will be raised from 50 per cent to 77 per cent. It will also be noted that a corresponding increase to 77 per cent in the number of satisfactory employees will be achieved by a test with a validity of only .50 if the selection ratio is decreased to .20.

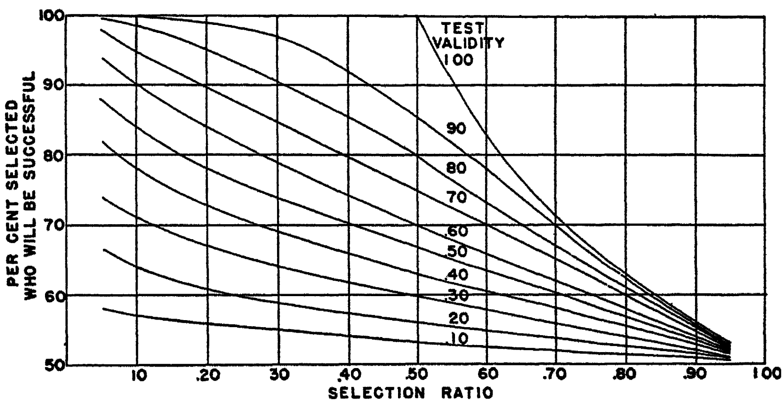


Fig. 5.5. Effect of test validity and the selection ratio upon the working efficiency of an employee selection test.

The Taylor-Russell tables,[13] reproduced in Appendix B, make it possible to determine what percentage of employees hired will be satisfactory under different combinations of test validity, selection ratio, and percentage of present employees considered satisfactory.

---

[13] Taylor and Russell, *op. cit.*

The use of these tables may be made clear by an example. Suppose an employment manager has a test with a validity coefficient of .40, has twice as many applicants or employees available as there are jobs to be filled, and is placing in a department where 30 per cent of the present employees are considered satisfactory. Looking in the upper half of the table on p. 649 (entitled "Proportion of present employees considered satisfactory = .30"), we find in the row representing a validity coefficient of .40 and in the column representing a selection ratio of .50, the value .41 where the indicated row and column cross. This means that under the conditions specified, 41 per cent of the employees placed will be satisfactory instead of the 30 per cent attained without the test. If conditions are such that the selection ratio may be still further reduced, the same test will place a still higher percentage of successful employees. For example, if only the highest 10 per cent of the persons tested are placed on the job, the percentage of satisfactory employees will be raised to 58 per cent, or nearly double the percentage of satisfactory employees placed without the test.

The "Direct Method" of Estimating Test Effectiveness In the above discussion, the use of the selection ratio requires that test validity be expressed in terms of a coefficient of correlation. McCollum and Savard [14] have shown that a simple and direct method can be used to obtain essentially the same results. With this method, which they call the Direct Method, the following steps are taken:

1. Select a cutting point on the test.
2. Select a cutting point on the criterion.
3. Count the number of cases falling above both cutting points: test and criterion.
4. Compare the number of cases found in (3) above with the number that would have been secured by random selection.

If 50 per cent of the entire group of employees are satisfactory, and if the test has no validity at all, 50 per cent of the employees above *any* so-called critical test score will be satisfactory. But if the test has *any* validity, more than 50 per cent of employees above a critical test score will be satisfactory. The extent to which the satisfactory percentage of those above the critical test score exceeds the satisfactory percentage without regard to the test score indicates the validity of the test in the particular situation.

McCollum and Savard make several empirical comparisons between this Direct Method and results obtained with the Taylor-Russell tables. The latter, of course, require the use of coefficients of correlation. Three

[14] I. N. McCollum and D. A. Savard, "A Simplified Method of Computing the Effectiveness of Tests in Selection," *Journal of Applied Psychology*, 41 (1957), 243–246.

such comparisons are summarized in Table 5.3, which is based on a table from McCollum and Savard's article.

Table 5.3

COMPARISON OF AVERAGE ERROR (IN PERCENTAGES) OF PREDICTION FOR TAYLOR-RUSSELL METHOD AND DIRECT METHOD FOR THREE GROUPS (50 AIRCRAFT WORKERS, 60 CLERICAL WORKERS, AND 46 MACHINE OPERATORS) OF INDUSTRIAL EMPLOYEES

				Average Error in Prediction		
Data Source	*r*	*Try-Out Group N*	*Follow-Up Group N*	*Taylor-Russell*	*Direct Method*	*Selection Test*
Air Craft Co.	.59	34	16	3.8	5.7	Wonderlic Personnel Test
Bellows	.57	30	30	9.0	2.2	Clerical Aptitude Test
Tiffin	.36	23	23	7.5	9.1	Bennett Test of Mechanical Comprehension
Average of all three groups				6.6	5.6	all above tests

McCollum and Savard, op. cit.

They divided the data from each study into a "Try-Out" group and a "Follow-Up" group. They then computed the validity coefficient for each "Try-Out" group and determined from the Taylor-Russell tables what per cent of satisfactory employees would be expected among subsequent groups of employees. The "Try-Out" groups were then subjected to the Direct Method analysis, and the expected percentages in this analysis were obtained.

Using selection ratios of .30, .40, .50, and .60 and proportions of satisfactory employees of .50, .60, .70 and .80, the average percentage error for all three studies, when the Taylor-Russell tables were used on the "Follow-Up" groups, was 6.6. The corresponding average error for the Direct Method was 5.6. These results strongly suggest that one can make effective use of the selection ratio concept without computing a coefficient of correlation.

Possible Limitations of Taylor-Russell Tables Several cautions concerning the use of the Taylor-Russell tables have been given by Smith.[15] The tabled values do not apply to triangular distributions of test scores plotted against a criterion. Figs. 5.3 and 5.4, which were used to explain the operation of the Taylor-Russell tables, assume that every increase in average test score is associated with an increase in average criterion measure. The tables further assume that the criterion measure in relation to test score is a linear function. Under certain conditions, neither of these assumptions is fulfilled. For example, it is sometimes found that success on a job increases with test scores up to a certain point, but that above this point, further increases in test scores bear no relation or even (in rare cases) a negative relation to job success. Guilford[16] states that an inspection of the scattergram between test scores and criterion is usually sufficient to determine whether the relation is essentially linear. If it seems desirable to test the linearity of the plot, several methods are available. If it is decided that the scattergram represents a definitely nonlinear function, the Taylor-Russell tables should not be used to predict the proportion of successful employees that will be obtained by using the test.[17]

While there is a definite theoretical point to the above caution advanced by Smith, some work by Tiffin and Vincent[18] suggests that it is usually safe to assume that the relationship shown by a Correlation Coefficient is sufficiently linear to justify the use of the Taylor-Russell tables.

Using 15 independent sets of predictor-criterion data, the theoretical expectancies obtained from modified[19] Taylor-Russell tables were compared with the empirical expectancies determined directly from the raw data. The data of each sample were split into fifths on the test score continuum, and the percentage of satisfactory employees in each test score category was computed directly. These percentages were then compared with the theoretical percentages predicted from the modified Taylor-Russell tables. In no one of the 15 sets of data did the empirical percentages differ from the tabled percentages more than could be accounted for by chance, and in the majority of cases there was rather remarkable agreement between the empirical and the theoretical expectancies. Fig.

---

[15] M. Smith, "Cautions Concerning the Use of the Taylor-Russell Tables in Employee Selection," *Journal of Applied Psychology*, 32 (1948), 595–600.

[16] J. P. Guilford, *Fundamental Statistics in Psychology and Education*, 3rd ed. (New York: McGraw-Hill Book Company), p. 149.

[17] *Ibid.*, p. 294.

[18] J. Tiffin and N. L. Vincent, "Comparisons of Empirical and Theoretical Expectancies," *Personnel Psychology*, 13 (1960), 59–64.

[19] Modified by C. H. Lawshe, R. A. Bolda, R. L. Brune, and G Auclair, "Expectancy Charts II. Their Theoretical Development," *Personal Psychology*, 11 (1958), 545–559.

5.6 gives an example of one of the worst fits from the 15 studies and Fig. 5.7 gives an example of one of the better fits. On the basis of these results, the authors concluded that in the majority of instances, it is advisable to use theoretical instead of empirical expectancies in constructing expectancy charts.

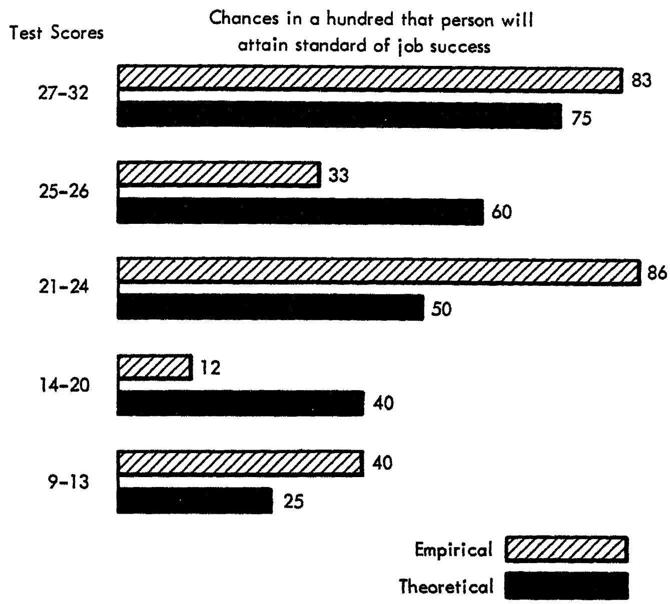


Fig. 5.6. An example of one of the poorest "fits" of empirical to theoretical expectancies. (See Tiffin and Vincent, *op. cit.*)

A second caution advanced by Smith concerns the source of the validity coefficient used with the Taylor-Russell tables. When, as is often the case, the validity coefficient of a test is based upon the present-employee method of "testing the test," the validity coefficient obtained will usually be lower than would have been obtained if the follow-up method had been used with a group of new employees. The reason for this difference is that many unsatisfactory employees have terminated their employment and only those doing at least well enough to remain on the job are available for the determination of the validity coefficient of the test. This reduction in the range of the criterion measures reduces the validity coefficient. The difference between the available and the true validity coefficients tends to result in underestimating the increase in satis-

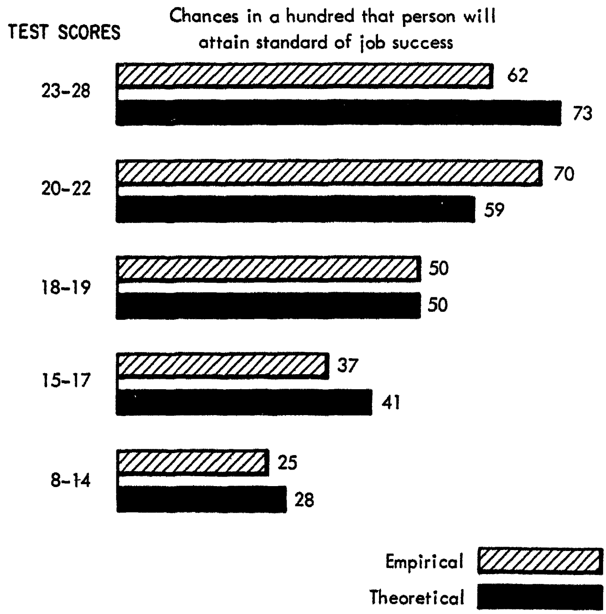


Fig. 5.7. An example of one of the better "fits" of empirical to theoretical expectancies. (See Tiffin and Vincent, *op. cit.*)

factory employees that would be obtained by using the test as a part of employment procedures.

An example of the application of the Taylor-Russell tables to a set of data is shown in Fig. 5.8. The validity coefficient of a battery of three

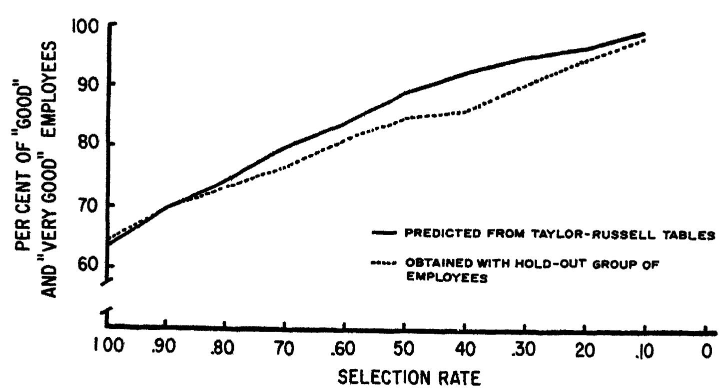


Fig. 5.8. Curves showing the per cents of high criterion employees predicted from the Taylor-Russell tables and the empirical results obtained with a "hold out" group of employees not used in determining the validity of the tests.

dexterity tests (treated finally as a single measuring instrument as discussed later on page 147) was found by Surgent[20] to be .76. A tryout of this battery on a "hold-out" group of employees, who had not been used in determining the validity coefficient, gave the results that have been plotted in Fig. 5.8. Interpolation from the Taylor-Russell tables gave expected theoretical percentages that also have been plotted in Fig. 5.8. The two curves plotted do not differ significantly. It is advisable whenever possible to make this kind of empirical check on the application of the Taylor-Russell tables to personnel test data.

INDIVIDUAL VS. INSTITUTIONAL EXPECTANCY CHARTS

The use of expectancy charts to show graphically the validity of a test was discussed on page 121, and illustrated in Fig. 5.1. There are two types of such charts—the individual and the institutional. The individual chart (which was illustrated in Fig. 5.1) shows what per cent of employees in *each test score bracket* will be superior on the job. Who is considered a "superior" employee (i.e., the criterion) must, of course, be decided before the analysis can be made. Another example of an individual expectancy chart is shown in Fig. 5.9. The employees tested in this study were maintenance men in an artificial ice plant. Each man was rated on a five point scale by the supervisor, and men receiving a rating of 4 or 5 were considered "superior." As Fig. 5.9 shows, if a man makes a test score between 103 and 120, there are 94 chances in 100 that he will

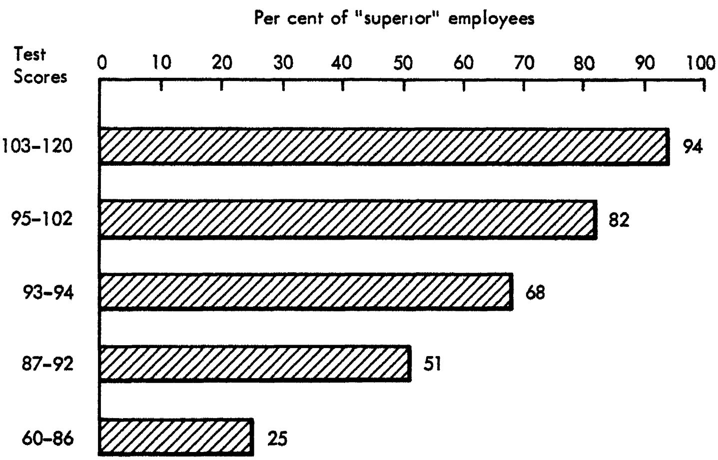


Fig. 5.9. Individual expectancy chart showing the per cent of "superior" employees as a function of score on the Purdue Mechanical Adaptability Test.

[20] L. V. Surgent, "The Use of Aptitude Tests in the Selection of Radio Tube Mounters," *Psychological Monographs,* 61 (No. 2, 1947), 1–40.

be given a rating that would classify him as superior. On the other hand, if he makes a test score between 60 and 86, he has only 25 chances in 100 of being rated superior. The individual expectancy chart thus permits individual prediction, i.e., the employing official can say what *each applicant's* chances are of being superior once he knows the applicant's test score on the test involved.

However, the primary task of the employment official is to be sure that, on the *average,* he is hiring men who will be most satisfactory on the job. He therefore will ordinarily be more concerned with the *institutional expectancy chart* than with the *individual expectancy chart.* The institutional chart shows the percentage of superior employees that will be obtained *if all applicants above a certain score are employed.* An example of an institutional chart constructed from the same data used in Fig. 5.9 is shown in Fig. 5.10.

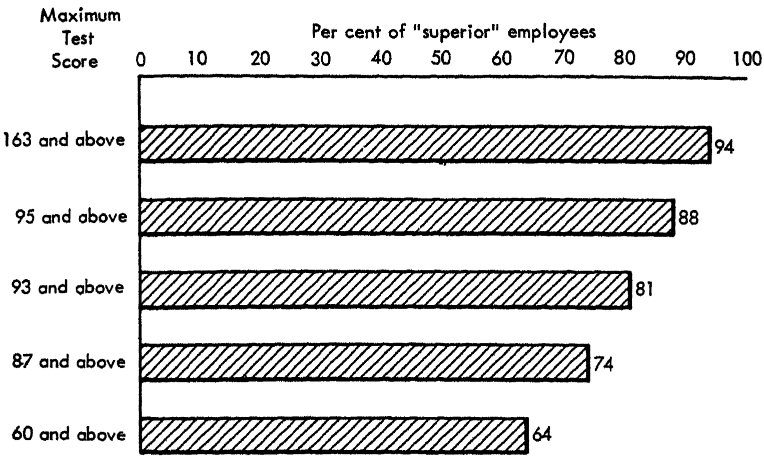


Fig. 5.10. Institutional expectancy chart showing the per cent of "superior" employees as a function of the *minimum* score required on the Purdue Mechanical Adaptability Test.

If the labor market will permit doing so, only men scoring 103 or above on the test would be employed. In this case, it can be expected that 94 out of 100 men employed, or 94 per cent, will be superior. If it is impossible to get enough men who make a score of 103, the hiring standard will need to be reduced. If it is necessary to employ men who score lower on the test, say down to 95, then only 88 per cent of those employed will be "superior." With an institutional expectancy chart available to him, the employment man can be constantly aware of the

chances he is taking as he reduces the hiring score when he is faced with a tight labor market.[21]

HOW TO USE TESTS FOR EMPLOYMENT

When a test has been found to be valid for a certain job, the next question to be answered is how the test should be used in the employment situation. What critical score (cut-off score) on the test should be required? Is it necessary, or desirable, to change the critical score from time to time? The answer to these questions is that the critical score should be as high as the labor market will permit. Since labor markets vary from one time to another, the critical score on the test likewise must vary.

The answers to these questions depend upon the use of test norms and the selection ratio under which the hiring is being done. An example in which both of these data were used will clarify the above comments. First, the validity of the Adaptability Test (see page 54) for first-line supervisors was determined. This investigation resulted in the institutional expectancy chart shown in Fig. 5.11. It is clear from Fig. 5.11 that there were 94 chances in 100 that men making scores of 18 or higher would survive as supervisors for a period of six months. Fig. 5.11 also

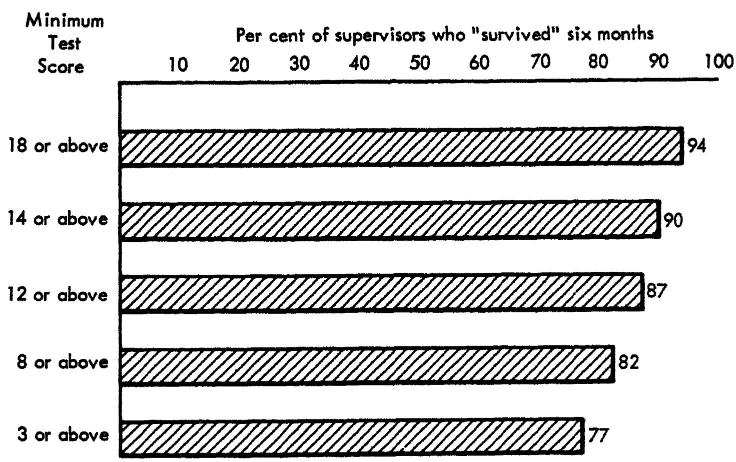


Fig. 5.11. Institutional expectancy chart showing the per cent of new supervisors who "survived" six months as a function of minimum score on the adaptability test.

[21] An excellent discussion of *individual* and *institutional* expectancy charts will be found in a book by L. E. Albright, J. R. Glennon, and W. J. Smith, *The Use of Psychological Tests in Business and Industry* (Cleveland, Ohio: Howard Allen, Inc., 1963).

shows the percentages surviving among men making successively lower critical scores. Obviously, in placing men in the supervisory job, the higher the Adaptability Test score, the greater are the chances that men will be selected who will be able to handle the new assignment. However, to decide without regard to the pool of eligible men that only men making a score of 18 or higher will be placed in the job is quite unrealistic. At this point, therefore, we should look at the norms on the test that have been based on the eligible men. These norms are shown in Table 5.4. They show the percentile ranks (see Appendix A, p. 615) equivalent to various raw scores for the group of men with whom we are concerned. As Table 5.4 shows, 100 per cent of the men made scores of 27 or below, 95 per cent made scores of 26 or below, 90 per cent made scores of 21 or below, 80 per cent made scores of 17 or below, and so on down the table. In other words, since 80 per cent of the men made scores of 17 or below, only 20 per cent made scores of 18 or above. This means that to find 20 men scoring 18 or above, we would need to have a pool of 100 men from which to select and we would therefore be using a selection ratio of .2. Since a pool of 100 men might not be available, we would have to reduce our critical score in order to fill the vacancies. If we needed 20 men and there were only 50 to choose among, we would be operating with a selection ratio of 20:50 or .4, which means that we would have to reduce the critical score to 13. The reason for this change in the critical score is that, as Table 5.4 shows, a score of 12 is at the 60th percentile, which means that 60 per cent of the group made a score of 12 or below, with only 40 per cent of the men scoring 13 or higher. Referring again to the Expectancy Chart shown in Fig. 5.11, it will be seen that we will not have as high a percentage of men surviving on the supervisor's job for six months when we use the critical score of 13 as we had when the critical score of 18 could be used.

Table 5.4

PERCENTILE NORMS ON THE ADAPTABILITY TEST FOR FOREMEN IN ONE PLANT

Percentile Rank	*Score on Test*
100	27
95	26
90	21
80	17
70	15
60	12
50	11
40	10
30	9
20	7
10	5
5	4
1	3

The point of the above discussion is that the critical (hiring) score on a test must be varied with the tightness or looseness of the labor market. The tighter the market, the lower the critical score. The looser the market, the higher the critical score can be.

COMBINING TESTS INTO A BATTERY

No single test will measure all of the capacities or abilities required on any job. Even the simplest of jobs is complex if one considers the combination of capacities or abilities required of a person who is to remain on the job and to do it well. The aptitude for any job consists of a syndrome of abilities, and one needs all of these to be successful. This fact makes it desirable, and in some cases necessary, to use a battery of tests rather than a single test. There are two basic methods of using tests in a battery—the *multiple cut-off* method and the *multiple correlation* method. With both methods, it is assumed that the validity of each test in the battery has been established for the job in question.

The multiple cut-off method [22] involves the application of the tests one at a time and elimination of applicants with each test who do not score at a satisfactory level. After the first test has been administered, certain low-scoring applicants are eliminated, and no further tests are administered to this group. In a similar way, after the second test has been administered, more applicants will be eliminated. This process is continued with all tests in the battery. After this has been completed, the only applicants remaining will be those who have made an acceptable minimum score on every test in the battery. Sometimes this method is used so that applicants (or candidates for promotion) are required to make a certain score on a stated number of tests in the battery, but do not have to "pass" all of the tests. In one investigation of this sort,[23] in which IBM proof machine operators were being hired, it was required that two of three tests be passed at the median score of the entire group tested. The battery used in this way had a very good selective efficiency and resulted in a 13 per cent increase in production.

The other method of combining tests requires the computation of a multiple correlation between the best combination of test scores and the criterion. By statistical methods [24] (beyond the scope of this book), the results of several tests can be combined into a composite score so that each is weighted to give the maximum correlation between the battery test score and the criterion. An example of the use of such combinations

---

[22] G. Grimsley, "A Comparative Study of the Wherry-Doolittle and a Multiple Cutting Score Method," *Psychological Monographs*, 63 (No. 2, 1949), 1–24.

[23] J. B. Harker, "Cross Validation of an IBM Proof Machine Battery," *Journal of Applied Psychology*, 44 (1960), 237–240.

[24] J. P. Guilford, *Fundamental Statistics in Psychology and Education*, 3rd ed. (New York: McGraw-Hill Book Company), pp. 411–415.

and weighting may be found in a set of tests worked out for placing menders in a hosiery mill. A number of tests were given to 100 employees on this job. For each employee, data were obtained on age and experience as well as average hourly earnings for the twelve-week period preceding the administration of the tests. The correlations between several of the tests and the earnings criterion are given in Table 5.5.

It will be noted in Table 5.5 that the maximum correlation of any individual test with earnings was .27, but that a battery made up of all three correlated .35 with the same criterion. In obtaining the composite score the following formula was used: Composite Test Score = 12 (Hayes Pegboard) − 4 (Purdue Hand Precision Test) − 2 (Finger Dexterity Error Score). The constants by which the raw test scores are multiplied in this formula result in the best combination of the tests to make a prediction of job success from a combination of test scores. The usefulness of a test battery that correlates to the extent of .35 with a criterion may be inferred from Fig. 5.12, which shows the percentage of employees

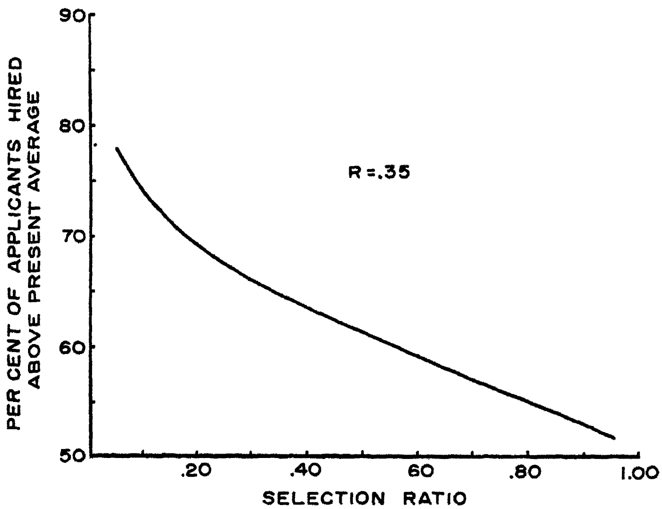


Fig. 5.12. Variation in per cent of employees who will be above the present average of employees when different selection ratios are used with a test having a validity coefficient of .35.

placed by this battery who will be above the average of present employees when different selection ratios are used in hiring or placement. For example, if the selection ratio could be reduced to .10, approximately 74 per cent of the employees placed would be above the present average. The value of the testing procedure in this case is definitely enhanced

by the use of several tests in combination, since no single test, of those tried, gave as high a validity coefficient as did the battery as a whole.

Table 5.5

CORRELATIONS BETWEEN SEVERAL DEXTERITY TESTS AND EARNINGS OF MENDERS IN A HOSIERY MILL

(Effect of age and experience statistically eliminated)

	Correlation with Earnings
Purdue Hand Precision Test	.27
Finger Dexterity Test (Error Score)	.18
Hayes Pegboard	.16
Composite Score of Three Above Tests	.35

In some cases the form of a scattergram of test scores and criterion values will itself indicate that a single test covers only one phase of the job requirements. This is especially indicated when a scattergram shows that persons scoring high on a test may be either high or low on the criterion, whereas virtually all of those testing low on the test are low on the criterion. An example of this is shown in Table 5.6. The scattergram in this table shows, for employees on a job in a hosiery mill, the relationship between scores on the Purdue Grooved Pegboard Dexterity Test (see p. 170), and a criterion of learning cost. (In this particular case the learning cost criterion was the amount the company had to "make up" between the actual piece rate earnings and the legal minimum wage.)

The triangular form of the scattergram shown in Table 5.6 indicates that some of the employees who had high test scores actually turned out to be slow learners (in the high learning-cost criterion groups), but

Table 5.6

SCATTERGRAM SHOWING RELATION BETWEEN LEARNING COST TO COMPANY OF 35 LOOPERS IN A HOSIERY MILL AND SCORES AT TIME OF EMPLOYMENT ON THE PURDUE GROOVED PEGBOARD DEXTERITY TEST

	Score on Grooved Pegboard Dexterity Test				
Learning Cost to Company	*60 seconds or over*	*55–59*	*50–54*	*45–49*	*40–44*
\$15–\$24			2	2	5
\$25–\$34		1	3	3	
\$35–\$44			2		3
\$45–\$59	1			3	
\$60 or over	2	3	3	1	1

none of those with poor test scores turned out to be rapid learners. The significance of this finding is that the test in question apparently is measuring one essential requirement of a rapid learner, but only one requirement. An employee who lacks this requirement is practically certain to be a slow learner. But an employee who tests high, that is, who has an abundance of this requirement, may still be a slow learner if he lacks certain other basic requirements for the job. The triangular shape of the scattergram thus indicates the existence of a hierarchy or syndrome of capacities required for this job.

For the reasons discussed on p. 140, the Taylor-Russell table predictions are not applicable to data of the type shown in Table 5.6. Although reducing the selection ratio in this situation will increase the proportion of rapid learners, the tabled expectancies will not be correct because the scattergram is triangular.

Analyses of combinations of tests in predicting job performance criteria have led to some interesting implications regarding interactions among tests. In this connection, Ghiselli [25] presents some evidence that indicates that the performance of *some* individuals can be predicted from tests more accurately than the performance of *other* individuals. But beyond this, he has found, in the case of certain jobs, that it is possible to use *one* test to identify those individuals for whom *another* test (or tests) *could* be used to predict their job performance. As he puts it,[26] it seems possible (in certain cases) that one can predict predictability! In a study of executives, for example, Ghiselli [27] reports that one test (referred to as a "moderator" variable or moderator test) was used to differentiate those individuals whose job performance could be adequately predicted by another test (a "predictor" test). For the entire group, the validity coefficient of the predictor test was .41. For the 21 per cent with lowest scores on the moderator test, however, the validity of the predictor test was only .10, whereas for the 26 per cent with highest scores on the moderator test, the validity of the predictor test was .68.

It should be noted, however, that the identification of the moderator and the predictor tests presumably needs to be done through research in the specific situation. Even though this interaction seems to be fairly situational, it suggests directions of research that may lead to improved prediction of job performance.

---

[25] E. E. Ghiselli, "Differentiation of Individuals in Terms of their Predictability," *Journal of Applied Psychology*, **40** (1956), 374–377.

[26] E. E. Ghiselli, "The Prediction of Predictability," *Educational and Psychological Measurement*, **20** (No. 1, 1960), 1–8.

[27] E. E. Ghiselli, "Moderating Effects and Differential Reliability and Validity," *Journal of Applied Psychology*, **47** (1963), 81–86.

ADEQUATE TRAINING OF TESTERS

The impression that the tester makes upon applicants or employees is of vital importance to the success of a testing program. This fact, obvious to the experienced employment manager, has often been underemphasized or even completely overlooked by psychologists themselves. "Everyone has shoes but the shoemaker's wife," says an old proverb. The professional psychologist often becomes so absorbed in the statistical phases of a test program that he forgets the importance of personal and industrial relations in actual test administration. Often the test administrator spends more time with an applicant than anyone else in the employment office. First impressions are lasting ones, and applicants who are employed may long remember the impressions of the company formed the first day at the plant. And the rejected applicants—upon whose reactions the industrial relations of the company within the community are also dependent—have no further opportunity to acquire a different opinion of the company. The man who administers the tests should be understanding, sympathetic, and courteous. He should give the applicant a feeling of importance. If the test results do not justify employment the applicant should be made to understand that, although he is not properly adapted or trained for the job now open, he may be quite qualified for some other job at a later time or for some job now open in another plant. The applicant should be given an understanding of the fact that he himself would not profit from employment on a job on which he would be likely to fail.

To explain such matters to rejected applicants is neither easy nor routine. It must be done individually, thoroughly, and sincerely. Many of the qualities of a successful salesman are of real value to the test administrator in industry. Testing human beings is not the same as testing materials or processes. People *react*, either favorably or unfavorably, to a test situation. It is one job of the tester—indeed, one of his most important jobs—to be sure that they react favorably. Others may have prepared adequate tests; light, roomy, and attractive rooms may be available in which to give the tests; but unless the test administrator creates a favorable reaction to the testing program, the full advantage of this tool for employee selection and placement is not likely to be realized.

6

Aptitude Tests

Tests that are used for prediction measure basic human characteristics or abilities that are related to the *capacity* to develop proficiency on specific jobs. These basic characteristics and abilities can be thought of as aptitudes. In this chapter we will be concerned with tests that measure aptitudes. Various aptitude tests will be mentioned or illustrated, along with a few examples of how aptitude tests are being used in personnel selection and placement.

KINDS OF APTITUDES

Psychologists have done a great deal of research in attempting to determine what the basic human characteristics are, and to develop tests to measure them. Much remains to be learned about such characteristics, but commendable progress has been made in their identification and measurement.

The characteristics we call "aptitudes" can be grouped into five classes, as follows: (1) mental abilities, (2) mechanical and related abilities, (3) psychomotor abilities, (4) visual skills, and (5) a class of other specialized aptitudes. It will be useful to discuss these classes separately.

MENTAL ABILITY TESTS

The term *mental ability* is used synonymously with the term *intelligence.* The field of mental ability testing has been more thoroughly explored than any other area of testing. As a result, we know the occupational fields in which mental ability tests are most likely to be of potential value. (The attention that mental ability testing has received, however, does not indicate that it is necessarily any more important than other types of testing.)

The Nature of Mental Abilities Although we frequently speak of mental *ability,* we should more properly use the expression mental *abilities,* since it has been found that there are various types of mental abilities. The most comprehensive studies of the nature of mental abilities have been carried out by Thurstone.[1] By a series of factor analyses [2] of mental tests he has identified various primary mental abilities of people. These include the following:

Verbal comprehension.
Word fluency.
Memory.
Inductive reasoning.
Number facility.
Speed of perception.
Spatial visualization.

Although the specific natures of some mental abilities are not yet clear, and though various studies of such abilities are not all in complete agreement, there is nonetheless strong evidence, such as that mentioned above, to suggest that there are several different kinds of mental abilities. These factors are correlated with each other, although the correlations among some of them are much higher than among others. The speed-of-perception and spatial-visualization factors will be discussed later, under "specialized aptitudes" and "mechanical ability" respectively.

Thurstone's studies suggest that each primary mental ability can be measured by a test specifically designed for that ability. Accordingly, such tests have been developed. On the other hand, there are some mental ability tests that include questions or items of different types, resulting therefore in a single score based on the composite of the types of items included in the test. Such tests typically include verbal, numerical, and reasoning questions. Tests of individual mental abilities have been found

---

1 L. L. Thurstone, *Primary Mental Abilities,* The Psychometric Laboratory, The University of Chicago, No. 50, September 1948.

2 *Factor analysis* is a statistical procedure used to identify basic "factors" in sets of data such as tests.

to be useful in many circumstances; and mental ability tests of the "omnibus" or composite nature have also demonstrated their practicality on a wide basis.

Although there are certain "individual" mental ability tests available, these are used mostly for counseling and clinical purposes. Most of the mental ability tests used for personnel purposes are group tests of a paper-and-pencil nature, so our interest is largely with such tests.

Typical Mental Ability Tests There are many different standardized mental ability tests that are available. A very comprehensive listing of such tests is in *The Fifth Mental Measurements Yearbook*.[3] For most tests, this *Yearbook* gives the title, a description of the group for which the test was constructed, the date of copyright or publication, whether the test is individual or group, the number of forms, the cost, the time required for administration, the author and the publisher, references to studies dealing with validity and reliability, and a brief evaluation of the test by one or more competent authorities. The industrial personnel man looking for a test that will serve some particular purpose can usually decide from the information given which test will be most likely to serve his purpose.

An additional source is *Tests in Print*.[4] This volume includes a listing of tests of many types used in education, psychology, and industry, and includes some tests that have been published since the *Yearbook*. It does not include descriptive and evaluative material, however, as does the *Yearbook*.

Appendix D of this book includes a description of a number of tests used in industry, along with information regarding their source. Some of these tests will be listed or mentioned in this, and subsequent, chapters. For further information about them the reader is referred to Appendix D.

In connection with mental ability tests, a few that are used in industry are listed below:

Otis Self-Administering Tests of Mental Ability
Wonderlic Personnel Test
The Adaptability Test
The Purdue Non-Language Personnel Test
SRA Primary Mental Abilities Tests

Examples of Mental Ability Tests in Use Since mental ability tests have been used with a wide variety of jobs, only a few typical examples of their use in personnel selection and placement will be given here.

---

[3] O. K. Buros, ed., *The Fifth Mental Measurements Yearbook* (Highland Park, N. J.: The Gryphon Press, 1959).

[4] O. K. Buros, *Tests in Print* (Highland Park, N. J.: The Gryphon Press, 1961).

Clerical jobs. This is one of the job areas in which mental ability tests have proved to be especially useful. It must be recognized, however, that clerical jobs vary greatly in their mental requirements. Where performance on a job depends predominantly upon other aptitudes or abilities, a mental ability test shows little correlation with job success.

Typical of studies in this field is one by Shore[5] with a sample of 93 bookkeeping-machine operators, 57 proof-machine operators, and 77 tellers in a large bank. These employees were administered several tests, including the Otis Self-Administering Test of Mental Ability and, in the case of the tellers, the Thurstone Test of Mental Alertness. The test scores were correlated with supervisory ratings. The correlations for most of the tests are given in Table 6.1. Since our interest here is in mental ability tests, the correlations for the Otis Test and, in the case of the

Table 6.1

CORRELATIONS * OF TEST SCORES WITH SUPERVISORY RATINGS FOR THREE GROUPS OF EMPLOYEES IN A LARGE BANK

Test	*Bookkeeping machine operators*	*Proof-machine operators*	*Tellers*
1. Otis Self-Administering Test of Mental Ability	.41	.15	.53
2. Short Employment Test			
Verbal	.44	.12	.48
Numerical	.48	.18	.62
Clerical Aptitude	.44	.29	.56
3. Minnesota Clerical Test			
Numbers	.44	.44	.45
Names	.38	.27	.40
Total Score	.39	.36	.40
4. MacQuarrie Test for Mechanical Ability			
Tracing	.14	.51	
Tapping	.21	.24	
Dotting	.00	.16	
Copying	.40	.13	
Location	.34	.23	
Blocks	.46	.33	
Pursuit	.46	.48	
Total Score	.48	.36	
5. Thurstone Test of Mental Alertness			
Linguistic			.63
Quantitative			.53
Total Score			.55

* *The correlations in the first and third columns are biserial correlations; those in the second column are Pearson correlations.*

[5] R. P. Shore, "Validity Information Exchange," Nos. 11–22, 11–23, and 11–24, *Personnel Psychology*, 11 (1958), 435–439.

tellers, the Thurstone Test of Mental Ability are of particular interest. The correlations for the other tests, however, will illustrate the usually desirable practice of validating several potentially relevant tests at the same time.

In another study by one of the present authors, the Adaptability Test was administered to clerical employees in several departments of a paper company. The correlation of this test with supervisors' ratings ranged from .40 to .65 for the employees in the various departments. In one department where only 57 per cent of the current employees were rated as "satisfactory," it was found that among those scoring 25 or above on the test, 86 per cent were "satisfactory."

Supervisors. Mental ability tests, in general, have "come through" more consistently in the selection of supervisors than any other type of test. As an illustration, a rubber company promoted 70 operators to supervisory jobs. To train these men for their new duties and responsibilities, the company conducted a series of supervisory training classes, at an expense to the company of $225 per man for the time spent in attending the classes. This amount did not include instructional costs or other expenses. During the first meeting of the supervisory training class, the men were given the Adaptability Test. The test scores were filed and a period of six months was allowed to pass while various criteria of success on the job of the new supervisors were investigated. At the end of the six-month period, it became apparent that the most satisfactory criterion of success available in this instance was whether or not the man was a good enough supervisor to be still on his new job six months later. Approximately one-fourth of the men were not, having quit or been demoted, transferred, or dismissed. In the case of the men no longer on the job, the cost of the supervisory training, as well as the other considerable expenses incurred in breaking in the new men, was borne by the company without return. Fig. 6.1 shows the percentage of men still on the supervisory job six months later among those scoring at various levels on the Adaptability Test.

As Fig. 6.1 shows, not a single man of those making scores of 4 or below on the Adaptability Test was still on the supervisory job six months after promotion. This company had selected the men for promotion primarily on the basis of their productivity on their previous nonsupervisory jobs. Promotion of those men scoring very low on the Adaptability Test turned out to be a serious mistake. They simply were not alert enough mentally to handle the supervisory job.

In addition to remaining longer on supervisory jobs, other studies have shown that men scoring relatively high on a mental ability test do better work on jobs of this type. Lawshe [6] reports a study in which repre-

[6] C. H. Lawshe, Jr., "How Can We Pick Better Supervisors?" *Personnel Psychology,* 2 (1949), 69–73.

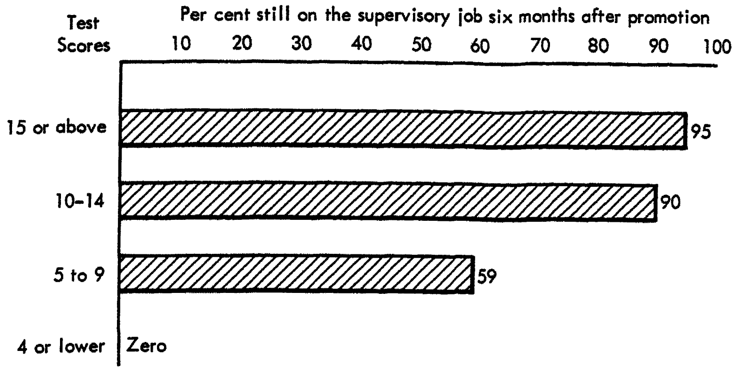


Fig. 6.1. Per cent of men promoted to supervisory jobs in a rubber company who were still on the job six months later as a function of score made on the adaptability test.

sentatives of 44 different companies, widely scattered geographically, each identified two of their "very best first-line supervisors" and two of their "poorest first-line supervisors." The four men in each plant, 176 in all, were given the Adaptability Test. The results of this study are shown in Fig. 6.2. Lawshe[7] interprets Fig. 6.2 as follows: "Note that when we

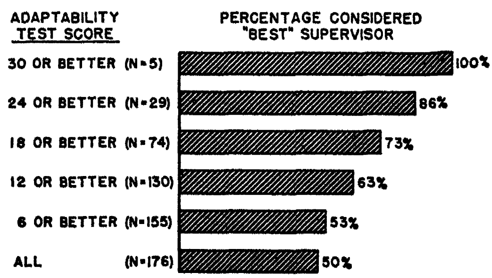


Fig. 6.2. The per cent of employees at or above various test score levels who were considered "best" in 44 industrial plants.

talk about all of these supervisors, 50 per cent are considered best, since each plant identified equal numbers of poor and best. It seems quite significant, however, that when we direct our attention to the 130 who did best on the test (those making a score of 12 or better) the percentage considered best is 63. Note that as successively higher minimum test scores

---

[7] *Ibid.*

are established, correspondingly higher percentages of supervisors appear in the best category."

In another investigation, Poe and Berg [8] administered the Adaptability Test to 33 steel mill production supervisors who had been rated in job performance by higher-level management. The ten highest-rated supervisors made an average score of 26.7 on the test. The ten lowest-rated men made an average score of only 18.1. The difference in average score was highly significant. In still another study, Dulsky and Krout [9] reported a correlation of .35 between Wonderlic Personnel Test scores and ratings by higher-level executives of 14 factory supervisors.

A further study on this topic has been reported by Bruce,[10] who administered the Otis Test of Mental Ability to 107 foremen in a tobacco company. The foremen were rated by the plant superintendent and the personnel manager. The combination of these three ratings was used as the criterion, and the Otis test scores correlated .29 with this composite rating.

In another test validation study dealing with 56 foremen in a manufacturing company,[11] one of the tests that was used experimentally was the Lorge-Thorndike Intelligence Test (Level 4), Verbal; this test correlated significantly ($r = .31$) with a criterion of paired-comparison ratings of the foremen that were made by their supervisors. The results of these and other studies have indicated quite consistently the fact that one of the primary requirements for supervisory jobs is mental ability. While there are of course also other personal characteristics that are related to supervisory performance (for example, personality characteristics), a mental ability test generally should be incorporated into a battery of tests for use in selection of individuals for supervisory positions.

Other types of jobs. Mental ability tests have also been found to be useful in the selection of personnel for various other types of jobs. For example, Maher [12] reports a correlation of .34 between scores on the Thurstone Test of Mental Alertness and performance ratings for 45 newspaper writers. In another situation, Giese [13] administered the Adaptability Test to the men on salaried office jobs of a midwestern manufacturing company. The men had been rated in over-all job performance by their

---

[8] W. A. Poe and I. A. Berg, "Psychological Test Performance of Steel Industry Production Supervisors," *Journal of Applied Psychology*, **36** (1952), 234–237.

[9] S. G. Dulsky and M. H Krout, "Predicting Promotion Potential on the Basis of Psychological Tests," *Personnel Psychology*, **3** (1950), 345–351.

[10] M. M. Bruce, "The Prediction of Success as a Factory Foreman," *Psychological Monographs*, **67** (No. 12, 1953), 1–17.

[11] J. A. Bromer, J. M. Johnson, and P. Sevransky, "Validity Information Exchange," No 15–02, *Personnel Psychology*, **15** (1962), 107–109.

[12] H. Maher, "Validity Information Exchange," No. 16–01, *Personnel Psychology*, **16** (1963), 71–73.

[13] W. J. Giese, "A Test Method for Selecting Office Personnel," *Personnel Psychology*, **2** (1949), 525–545.

supervisors. Giese found that of men scoring 14 or above on the Adaptability Test, 91 per cent had been rated above average on the job. Of those scoring 13 or below, only 9 per cent were rated above average.

Another area in which the use of a mental ability test has been found very helpful is in the selection of teletype operator trainees. It is customary for the operating company to send trainees for this job to a special school where a carefully prepared program of instruction is given. The trainees are graded periodically by means of standardized and objective tests that cover both job performance and related information. Trainees are certified as operators only after successfully passing this course.

In one investigation, 219 teletype trainees were given the Adaptability Test, previously described, shortly after being selected for training. The follow-up method of test validation was employed to determine the relation between test scores and success in the training course. The results are summarized in Table 6.2.

Table 6.2

RELATION BETWEEN SCORES ON THE ADAPTABILITY TEST AND SUCCESS IN A TRAINING COURSE FOR TELETYPE OPERATORS

Score on Adaptability Test	*Number Passing Course*	*Number Failing Course*	*Total Number*	*Per Cent Passing Course*
22 or above	27	1	28	96%
16–21	66	4	70	94%
10–15	83	10	93	89%
9 or below	17	11	28	61%

The cost to the company for each trainee (which includes transportation to and from the school, board and room, salary during that training, and the pro-rata cost per trainee of operating the school) amounts to approximately $500 per trainee. The investigation cited above shows that from every 100 trainees who score 22 or above on the Adaptability Test, the company will be able to certify 96 operators, the cost thus being $\frac{100 \times \$500}{96}$ or $520.83 per operator. Among trainees scoring 9 or below on the Adaptability Test, of every 100 trained only 61 will be certified as operators, the cost thus being $\frac{100 \times \$500}{61}$ or $819.67 per operator. The comparative costs show that *each of the operators in the low-score bracket costs the company $298.84, or approximately 57 per cent, more in training expenses than do operators in the high-score bracket.*

There are, of course, many jobs for which mental ability tests have

been found *not* to be useful in selection. In such cases, job performance presumably depends upon other aptitudes or abilities, and the degree of mental ability of individual workers neither increases nor decreases the chances of job success.

Still other investigations have shown that mental ability tests sometimes show a negative relationship with success on the job. For example, in one study of routine assembly workers, it was found that there was a negative correlation between scores in the Otis Self-Administering Test of Mental Ability and criteria of actual production and of supervisors' ratings of job performance. And in another situation, it was found that the best inspectors of simple material had lower average scores on the Adaptability Test than did the poorest inspectors on the same jobs. These studies are in accord with the very early studies by Bills [14] and by Pond and Bills [15] showing that individuals who test high on mental ability usually should not be placed on routine, easy jobs, and that frequently persons scoring very low on mental ability tests consistently do better on simple, repetitive jobs than do persons whose mental ability is average or above.

MECHANICAL ABILITY AND RELATED TESTS

The term "mechanical ability" has been rather loosely used to characterize a wide range of jobs that involve working with mechanical parts and equipment. We should hasten to say that essentially two *types* of abilities are required, in varying degrees, in "mechanical" jobs. There are certain *mental* aspects, such as mechanical comprehension and the understanding of mechanical principles; and, there are *motor* or *physical* skills such as muscular coordination and dexterity. We will consider mechanical ability in the sense in which it is used by Thurstone [16] as "a complex of intellectual abilities." Although it is true that mechanics, craftsmen, and skilled machine operators do require certain motor skills in order to perform their jobs, such employees usually succeed or fail in proportion to their training and general mechanical comprehension, not in proportion to their muscular skills. Mechanical ability tests are used primarily to select employees for jobs that require a mechanical "knack," such as in the maintenance of machinery, in "setting up" production machinery before it is turned over to an operator, or in the repair of household appliances.

---

14 M. A. Bills, "Relation of Mental Alertness Test Scores to Positions and Permanency in Company," *Journal of Applied Psychology*, 7 (1923), 154–156.

15 M. Pond and M. A. Bills, "Intelligence and Clerical Jobs. Two Studies of Relation of Test Score to Job Held," *Personnel Journal*, 12 (1933), 41–43.

16 L. L. Thurstone, "An Analysis of Mechanical Aptitude," *The Psychometric Laboratory*, The University of Chicago, No. 62, January, 1951.

The Nature of Mechanical Ability Although we are considering mechanical ability in the restricted sense of mental ability, we still want to know whether there is a single *general* mechanical ability or whether there are *various* such abilities. The accumulating evidence of psychological research seems to justify the conclusion that there are various *kinds* of abilities that may appropriately be considered as mechanical abilities. There is, however, a central or principal mechanical ability that we can consider as something of a general human ability. This was shown, for example, in factor analyses by Friedman and Ivens[17] and Friedman and Detter[18] of certain Air Force and civilian tests. In both of these studies a factor called "mechanical experience" was identified, this representing the knowledge gained from experience with mechanical apparatus and tools. For convenience of terminology, we might call this "general mechanical aptitude."

In addition to this general factor, however, other more specific factors were identified which appropriately may be considered as falling within the area of mechanical abilities. Both studies revealed a factor of *visualization;* this represents the ability to manipulate mentally visual images in two or three dimensions, as in visualizing the forms that various geometric figures would make. This presumably corresponds to the spatial visualization factor reported by Thurstone in his studies of primary mental abilities previously discussed under mental ability tests.

Another factor reported by Friedman and Ivens[19] was one called *spatial relations.* This may appear to be somewhat akin to the visualization factor, but there is a difference. This factor represents the ability to *perceive* spatial patterns accurately, and to *compare* them with each other, such as in recognizing objects to be identical or different. The visualization factor deals more with the ability to manipulate visual images. It should be added that there is some evidence that there may be other spatial factors in addition to the ones mentioned above.

Although the "general mechanical aptitude" factor probably is the most important in many mechanical jobs, it can readily be seen that these factors of visualization and spatial relations would be of considerable importance in some mechanical jobs, since some such jobs require the ability to visualize the physical relationships of objects, or to identify objects as being similar or different. It should also be pointed

---

[17] Gabriel Friedman and Frank C. Ivens, "Factor Analysis of the Airman Classification Battery AC—1B; the USES General Aptitude Test Battery, Experimental Visualization and Spatial Tests, and Psychomotor Tests," Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas, *Research Bulletin,* AFPTRC–TR–54–67.

[18] Gabriel Friedman and Howard M. Detter, "Factor Analysis of Airman Classification Battery and Selected Air Force and Civilian Tests," *Research Bulletin,* AFPTRC–TR–54–75.

[19] Friedman and Ivens, *op. cit.*

out that these various factors are not entirely independent, but tend to be correlated with each other.

Tests of Mechanical Ability A few of the most commonly used tests of mechanical ability are listed below:

Group tests
- Bennett Test of Mechanical Comprehension
- Flanagan Aptitude Classification
  Tests: Mechanics (No. 13) and Assembly (No. 5)
- Purdue Mechanical Adaptability Test
- Purdue Industrial Training Classification Test
- Revised Minnesota Paper Form Board

Individual Tests
- Minnesota Mechanical Assembly Test
- Minnesota Spatial Relations Test
- Purdue Mechanical Performance Test

These, and other, such tests are described or listed in Appendix D, along with their sources. Examples of certain of these, or items from them, are shown in Figs. 6.3, 6.4, 6.5, and 6.6. In addition, particular mention will be made of the Purdue Mechanical Adaptability Test, to illustrate the processes that are involved in developing standardized tests of this type.

The Purdue Mechanical Adaptability Test was developed to aid in identifying men or boys who are mechanically inclined, and who, therefore, are most likely to succeed on jobs or in training programs calling for mechanical abilities and interests. The test measures one's experiential background in mechanical, electrical, and related activities. The test was constructed to measure experiential background because there was reason to believe from a previous study that, other things being equal, those persons who have most profited in knowledge from previous mechanical experiences may do better on mechanical jobs than those persons who have not so profited. The test is one of general mechanical aptitude.

The questions comprising Form A of the Purdue Mechanical Adaptability Test were selected by statistical methods designed to achieve maximum reliability of the final test and as low a correlation as possible with general intelligence. The odd-*vs.*-even items method of determining the reliability, using a group of 487 industrial applicants, resulted in a reliability coefficient of .84. With the same group of applicants, the correlation between scores on the Purdue Mechanical Adaptability Test and the Adaptability Test was found to be .32. With another group of 329 employees of a plumbing manufacturing company, the correlation between Mechanical Adaptability Test scores and Adaptability Test scores was found to be only .18. A later study of 382 employees of an electrical manufacturing company showed a correlation of .26 between scores on the Purdue Mechanical Adaptability Test and the Wonderlic Personnel Test.

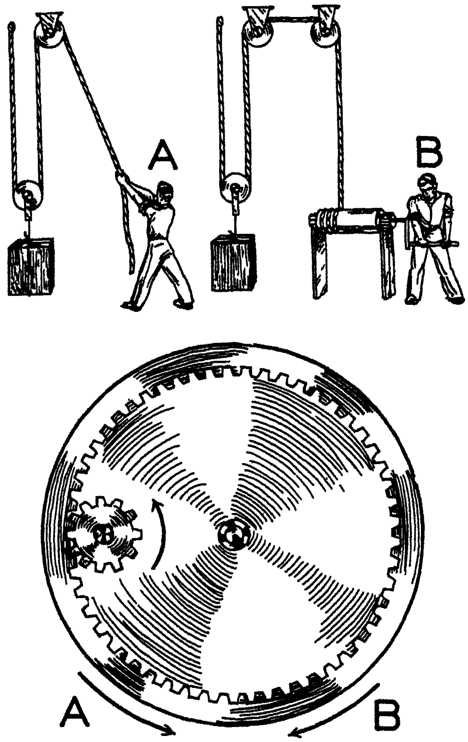


Fig. 6.3. Two items from Form AA of the Bennett Test of Mechanical Comprehension.

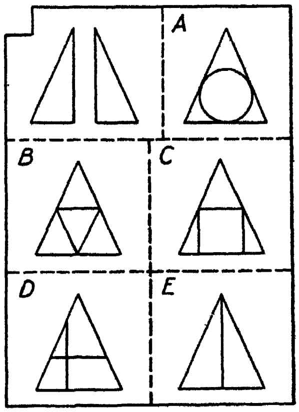


Fig. 6.4. An item from the Minnesota Paper Form Board Test. The person tested is asked to select the set of lettered parts (A, B, C, D, or E) which may be formed by the parts shown in the upper-left square.

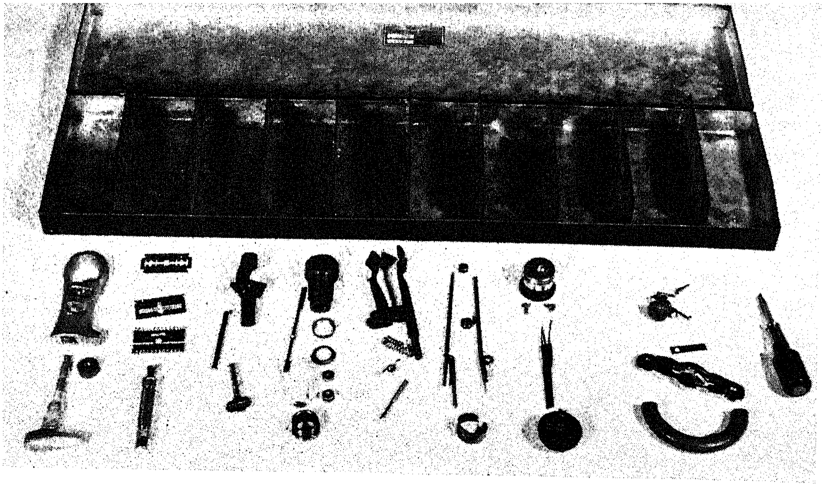
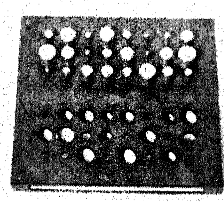
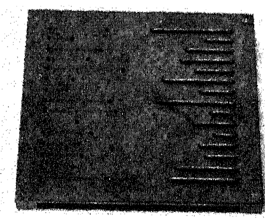


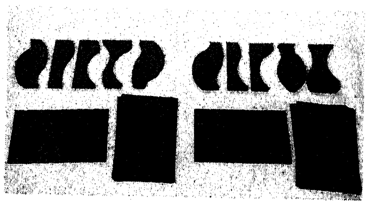
Fig. 6.5. One form of the Minnesota Assembly Test.



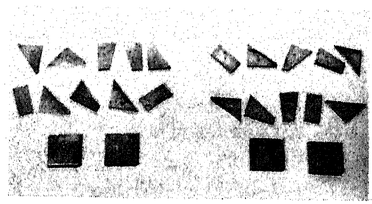
1. Transfer Board A.



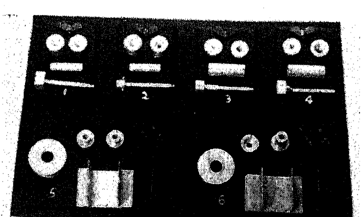
2. Transfer Board B.



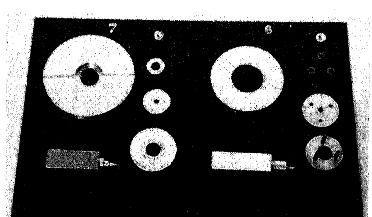
3. Cue Boards A and B.



4. Spatial Relations Blocks A and B.



5. Hub Assemblies 1–6.



6. Hub Assemblies 7 and 8.

Fig. 6.6. Purdue Mechanical Performance Test. (Distributed by the Lafayette Instrument Co., Lafayette, Ind.)

These correlational studies clearly show that the Purdue Mechanical Adaptability Test is *not* just another intelligence test.

Examples of Mechanical Ability Tests in Use . General mechanical ability and visualization and spatial relations abilities have generally been found to be requirements for jobs that involve an understanding of physical and mechanical principles, and the perception of the relationships of physical objects and their movements. Among the kinds of jobs for which these abilities seem to be especially relevant are apprentices, craftsmen, mechanics and repairmen, draftsmen, and engineers. However, there are also many other types of jobs for which such abilities are relevant. A few examples of test validation studies with tests of these abilities will be mentioned briefly. In one study, the Purdue Mechanical Adaptability Test was given to 246 railroad-trade apprentices. The apprentices were rated by their instructors and trainers. Those rated as "good" were considered to be performing satisfactory work. The relationship between the scores on the test, and the ratings, is shown in Fig. 6.7, which indicates that more of the apprentices with high test scores were "good" than were those with low test scores.

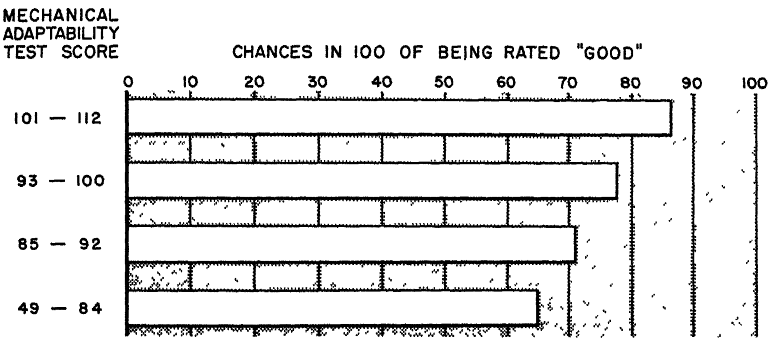


Fig. 6.7. Expectancy chart showing relation between Purdue Mechanical Adaptability Test Scores and rated success of railroad trade apprentices.

In another test validation analysis, the Purdue Mechanical Adaptability Test was given to a group of 34 airplane engine mechanic trainees. The men were rated by their instructors using the paired comparison method (see p. 233). The relationship between the test scores and ratings is shown in Fig. 6.8. Fig. 6.8 shows that there is a marked relation between the score made on the test and the trainee's chances of being rated above average.

Campbell[20] reports a study directed toward identifying tests that

[20] J. T. Campbell, "Validity Information Exchange," No. 16-04, *Personnel Psychology*, **16** (1963), 181–183.

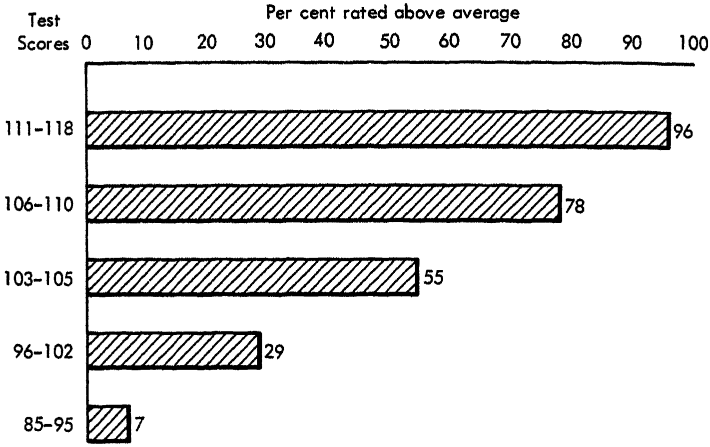


Fig. 6.8. Relation between scores on the Purdue Mechanical Adaptability Test and ratings of 34 airplane engine mechanic trainees.

would be useful in the selection of gas deliverymen. (This job involves the delivery and installation of gas tanks at customers' homes, adjusting the equipment and keeping it in operating condition, as well as serving as a contact with customers.) Of the three tests that were administered to the 155 men, the Bennett Test of Mechanical Comprehension was found to have the highest correlation with job performance (as rated by supervisors).

In still another study using the Purdue Mechanical Adaptability Test, 46 machine operators from a screw manufacturing company were rated by supervision for their proficiency on the job. A chart showing the percentage of superior operators receiving superior ratings among groups scoring successively higher on the Purdue Mechanical Adaptability Test is shown in Fig. 6.9. This chart indicates that of men scoring 60 or more (a group that included all the operators), only 37 per cent were rated superior, while among operators scoring 95 or above, 65 per cent were rated superior operators.

Owens[21] describes a form of the Bennett Test of Mechanical Comprehension (Form CC) that is more difficult than the previous forms. This test, when used with engineering students, "showed a satisfactory correlation with relevant course grades, was an excellent predictor of academic mortality, predicted better at high achievement than at low, made a significant independent contribution beyond that made by the

---

[21] W. A. Owens, Jr., "A Difficult New Test of Mechanical Comprehension," *Journal of Applied Psychology,* 34 (1950), 77–81. Reproduced by permission of the American Psychological Association.

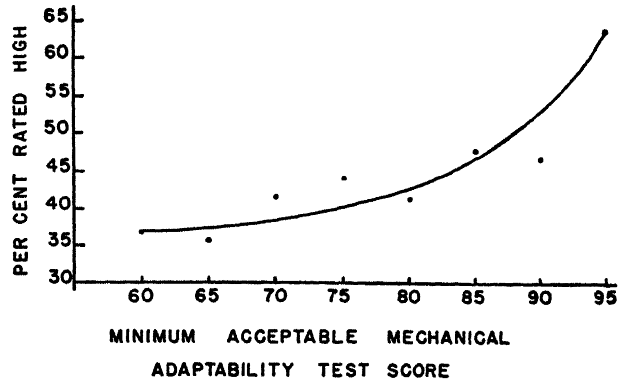


Fig. 6.9. Percentage of automatic screw machine operators who are rated "high" when successively higher test scores on the Purdue Mechanical Adaptability Test are used.

American Council on Education Psychological Examination and high-school average, and gave scores which depended on training or experience only to the extent of 2 or 3 score points.

In another investigation with engineers, Spencer and Reynolds[22] gave the Minnesota Engineering Analogues Test (MEAT) to 28 recent bachelor's degree mechanical and electrical engineering graduates who had been employed by a manufacturer of business machines. Scores on this test were correlated with three criteria as listed below; the correlations are also given:

Criterion	*Correlation*
1. Training program class standing (at 6 mo.)	.40
2. Job performance ratings (at 6 mo.)	.58
3. Job performance ratings (at 18 mo.)	.64

It is interesting to note that the validity of the test was higher at the end of 18 months of employment than at the end of the first 6 months.

The last example of mechanical ability tests we shall mention concerns engineering draftsman trainees, as reported by Ruch.[23] In this case, 133 high school graduates were administered a battery of seven tests, namely, those of the Employee Aptitude Survey. (One of these tests was a test of space visualization.) Of these applicants, 30 with the lowest scores were rejected immediately. The others were referred to the drafting training instructor, who selected 37 for the training program (but not on

[22] G M. Spencer and H. J. Reynolds, "Validity Information Exchange," No. 14-04, *Personnel Psychology*, 14 (1961), 456–458.

[23] F. L. Ruch, "Validity Information Exchange," No. 13-02, *Personnel Psychology*, 13 (1960), 448.

the basis of the tests). The following correlations were computed between scores on the tests and later supervisory ratings of job performance:

Test	*Correlation*
Verbal Comprehension	08
Numerical Ability	.51
Visual Speed and Accuracy	.45
Space Visualization	36
Numerical Reasoning	46
Verbal Reasoning	.63

While spatial visualization is apparently an important aspect of performance, it can be seen that other factors also are significantly related to performance on the job.

PSYCHOMOTOR TESTS

Psychomotor tests are those which measure muscular abilities, or combinations of sensory and muscular abilities. The term *psychomotor* covers the range of abilities that includes what we commonly call dexterity, manipulative ability, motor ability, and eye-hand coordination, as well as other aspects of muscular performance.

The Nature of Psychomotor Abilities There is quite a bit of support for saying that there are various *kinds* of psychomotor abilities. This has been supported by the results of various studies, such as by Friedman and Ivens,[24] and, in particular, those by Fleishman,[25] Fleishman and Hemphill,[26] and Fleishman and Ornstein.[27] Through a series of factor analyses and related studies, a number of psychomotor ability factors have been identified. Some of the more important of these are described below: [28]

Control precision. This factor involves fine, relatively controlled muscular adjustments, primarily where larger muscle groups are involved; it covers both arm-hand and leg movements. It is especially important where careful positioning of controls by the hands or feet is required.

Multilimb coordination. This factor is characterized by the simultaneous coördination of movements of the hands and/or the feet.

---

24 Friedman and Ivens, *op. cit.*

25 E. A. Fleishman, "Dimensional Analysis of Psychomotor Abilities," *Journal of Experimental Psychology*, 48 (1954), 437–454.

E. A. Fleishman, *The Description and Prediction of Perceptual-Motor Skill Learning*, paper presented at Symposium on Training Research and its Implications for Education, February 1–3, 1960, University of Pittsburgh (Sponsored by the Office of Naval Research).

26 E. A. Fleishman and W. E Hemphill, "Factorial Analysis of Complex Psychomotor Performance and Related Skills," *Journal of Applied Psychology*, 40 (1956), 96–104.

27 E. A. Fleishman and G. N. Ornstein, "An Analysis of Pilot Flying Performance in Terms of Component Abilities," *Journal of Applied Psychology*, 44 (1960), 146–155.

28 Fleishman (1960), *op. cit.*

Response orientation. This factor involves the ability to make the correct directional movement in relation to the correct stimulus, under highly speeded conditions.

Reaction time. This is simply the speed with which an individual is able to respond to a stimulus when it appears.

Speed of arm movement. This represents the speed with which an individual can make a gross, discrete arm movement where accuracy is not a requirement.

Rate control. This factor involves the making of continuous anticipatory motor adjustments relative to changes in speed and direction of a continuously moving target or object, as in a tracking task.

Manual dexterity. This ability involves skillful, well directed arm-hand movements in manipulating fairly large objects under speed conditions.

Finger dexterity. This is the ability to make skillful, controlled manipulations of tiny objects involving, primarily, the fingers.

Arm-hand steadiness. This is the ability to make precise arm-hand positioning movements where strength and speed are minimized.

Wrist-finger speed. This is essentially a "tapping" type of ability, with an emphasis on speed of tapping.

Aiming. This is a highly restricted type of psychomotor skill dealing specifically with the ability to aim the fingers successively and rapidly at a series of predetermined locations.

While psychomotor abilities generally can be considered to be relatively separate types of abilities, there is still some question as to the extent to which they may be considered as being *completely* independent, as opposed to being related to each other. In discussing this, however, Vernon[29] points out that in several studies with psychomotor tests, it has been found that such tests typically have some relationships with each other, although the average inter-correlations among groups of them are quite low, illustrations being .25, .27, .30, and .13. Thus, although such tests typically are correlated with each other to some degree, the correlations are so low that psychomotor abilities must be considered as being predominantly specific.

The reader might then raise this question: What are the practical implications of this in using psychomotor tests? The reply would be to the effect that it is *particularly* important in using psychomotor tests to be sure that the test or tests used in a given situation be thoroughly validated for the job in question before they are used. Performance on any one test *cannot* be assumed to give a measure of general psychomotor ability. It is necessary to use specific tests that are *known* to be related to job performance.

[29] P. E. Vernon, "The Structure of Human Abilities" (London: Methuen and Co., Ltd., and New York: John Wiley & Sons, Inc., 1950).

Aside from being relatively independent among themselves, psychomotor abilities also are relatively independent of other abilities and characteristics. Let us consider mental abilities, for example. Although the correlations between psychomotor and mental ability tests are somewhat variable (depending on the specific tests and samples of people), the correlations tend to be low, or zero. For example, in an investigation by the Western Electric Company,[30] the correlation between the O'Connor Finger Dexterity (described below) and the Otis Mental Ability Test was found to be .07. The correlation between the Hayes Pegboard (another dexterity test mentioned below) and the Otis Test was found to be zero. Each of these correlations was based on test scores of 749 employees. These results are typical of the results of many similar investigations. Although the correlations generally tend to be positive rather than zero, they typically are so low as to be of little practical value.

Somewhat the same can be said about the relationship between psychomotor tests and mechanical ability tests. In the study by Friedman and Ivens,[31] for example, the following correlations were reported for the specified psychomotor tests with a test of Mechanical Principles: Circle Dotting, .02; Two-Hand Drawing, .03; Two-Hand Coordination, .29; Rotary Pursuit, .19; and Finger Dexterity, .11. These very low correlations justify the conclusion that, for practical purposes, psychomotor abilities are relatively independent of mechanical abilities. Neither is there any appreciable relation between psychomotor abilities and anthropometric measurements, such as length of fingers or other body measurements.

Typical Psychomotor Tests Among the psychomotor tests that are used most commonly in personnel selection are the following:

O'Connor Finger Dexterity Test
O'Connor Tweezer Dexterity Test
Hayes Pegboard
Purdue Pegboard
Purdue Grooved Pegboard
Minnesota Rate of Manipulation Test
MacQuarrie Test for Mechanical Ability
Stromberg Dexterity Test

Some of these, and other, tests are listed in Appendix D. Illustrations of some of these are shown in Figs. 6.10, 6.11, 6.12, and 6.13.

Examples of Psychomotor Tests in Use A typical type of job for which psychomotor abilities are of considerable importance is that of

---

[30] *Analysis of 1935–37 Experience in Selecting New Men for Shop Occupations.* Privately printed Monograph (Chicago: Western Electric Company, Hawthorne Plant, 1939).

[31] Friedman and Ivens, *op. cit.*

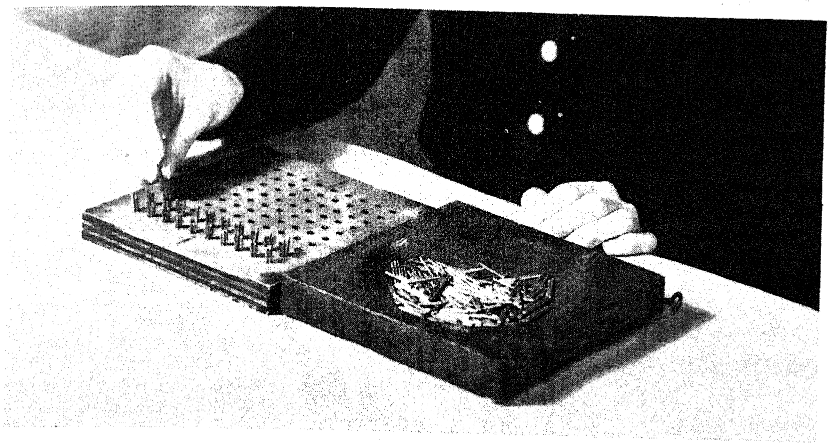


Fig. 6.10. O'Connor Finger Dexterity Test.

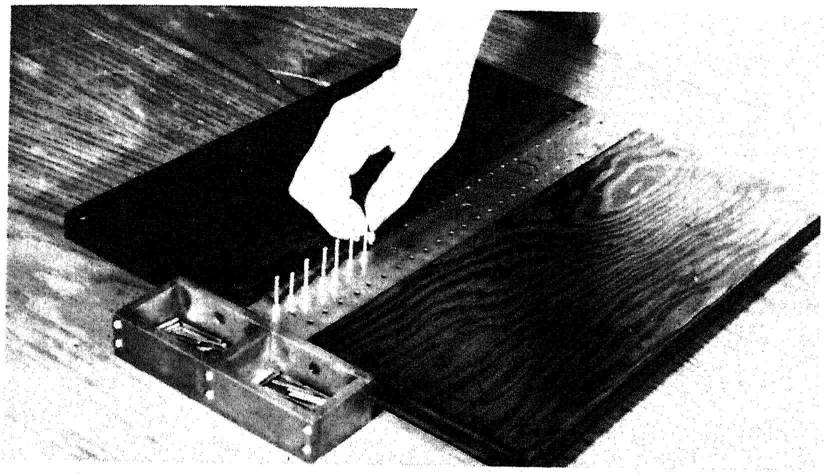


Fig. 6.11. The Hayes Pegboard.

packer. In a test validation study of packers in a handcraft company, Wolins and MacKinney [32] gave the Purdue Pegboard and the Minnesota Rate of Manipulation Test to 27 employees. (The job involves packing in boxes or envelopes handcraft kits containing parts to be assembled by customers.) The packers were rated by their supervisor, classifying each

[32] L. Wolins and A. C. MacKinney, "Validity Information Exchange," No. 15-04, *Personnel Psychology*, **15** (1961), 227–229.

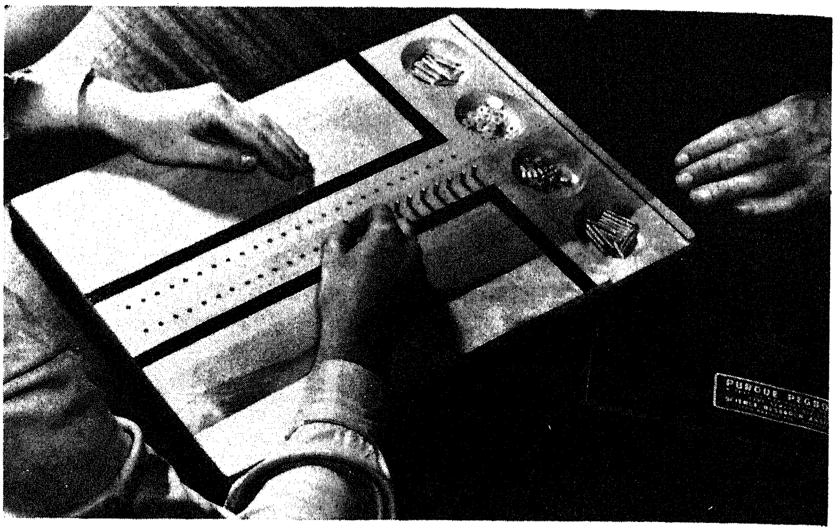
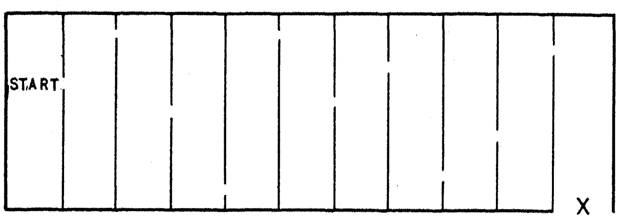
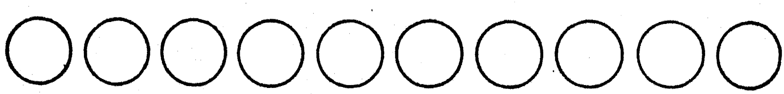


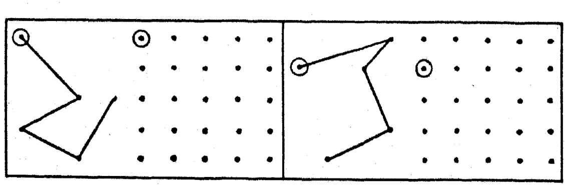
Fig. 6.12. Purdue Pegboard.



Tracing: Applicant draws a line from "start" to X without crossing a vertical line.



Dotting: Applicant places three dots in each circle as quickly as possible.



Copying: Applicant reproduces the drawings by connecting the appropriate dots in the area to the right of each drawing.

Fig. 6.13. Several parts of the MacQuarrie Test for Mechanical Ability. This test is essentially a group test of muscular co-ordination and control.

in one of five rating categories. The correlations of the various parts of the two tests with these ratings are given below:

Test	*Correlation* *
Purdue Pegboard	
Right Hand	.19
Left Hand	.32
Both Hands	.19
Right + Left + Both	.27
Assembly	.47
Minnesota Rate of Manipulation Test	
Placing	.21
Turning	.40
Displacing	.16
1 hand turn and place	.26
2 hand turn and place	.27

* The correlations for the Minnesota test were actually negative, since it is a work-limit test with scores being based on *time* to complete the test. Since good performance on the test (low time scores) is related to good job performance (and vice versa), the correlations are shown here as positive.

It can be seen from these data that all parts of both tests were correlated with the ratings, although certain parts had noticeably higher correlations than the others, especially the Assembly Part of the Purdue test, and the Turning Part of the Minnesota test.

As another example of the use of psychomotor tests, we will report the results of a study by Durrett [33] with continuous miner operators in a bituminous coal company. Three tests were administered to the men, and supervisory ratings were used as criteria. The correlations of the tests with the ratings are given below:

Test	*Correlation*
Revised Beta Examination	.00
Bennett Test of Mechanical Comprehension	.12
Two-Handed Coordination Test	.57

It can be seen that only the Two-Handed Coordination Test had any appreciable relationship with the criterion—but its correlation (.57) was a substantial one, indicating that the psychomotor coordination measured by it is an important job requirement.

Synthetic dexterity tests. Previous mention has been made of the concept of synthetic validity. In line with this concept, Drewes [34] developed an assortment of manipulative test materials (especially pegboard components) that could be used in various ways. After an analysis of the elemental motions of each of several jobs, he developed (using the test

[33] H. L. Durrett, "Validity Information Exchange," No. 14-03, *Personnel Psychology,* 14 (1961), 453–455.

[34] D. W. Drewes, "Development and Validation of Synthetic Dexterity Tests Based on Elemental Motion Analysis," *Journal of Applied Psychology,* **45** (1961), 179–185.

materials) a test for each job that approximated the combination of elemental motions in the job. These specially-designed tests generally were found to be more valid for prediction of job success than standardized tests that were used.

VISION TESTS

Practically every industrial job requires some degree of vision, and many jobs require a high degree of skill in some particular visual function. The inspector of small parts for "appearance" must have keen vision at close distances. The operator of certain knitting and other textile machines must not only have keen vision at close distances but also must be able to maintain such vision for long periods with only occasional interruption. The truck driver, crane operator, and signal man must have keen vision for greater distances and good perception of space relationships. Color discrimination is of importance to an employee wiring a radio (because he must discriminate between wires of different color), to the operator of a color-printing press, to the pipefitter tracing a color code through the plumbing system of a plant, and to the operator of mobile equipment who must depend upon colored signals to determine whether roads are open or pathways clear. Various measurable visual characteristics have been found to be related to successful performance on certain jobs—even on jobs in which these visual factors could not have been inferred from ordinary job analysis procedures. The pattern of visual requirements differs markedly for different jobs, and these requirements cannot be discovered except by a study of the people on the job, as explained in Chapter 5.

Management in production industries has long recognized the importance of good vision in employees.[35] Some form of vision test, administered either in the employment office or as a part of the medical examination, is perhaps more common in industry than is any other form of employee test. Testing visual skills has always been a legitimate psychological function along with testing carried on in other areas of skill or aptitude.

Visual skills are not innate and permanent; they are to a large degree acquired and can be, and often are, modified. They change universally with age. Professional eye care, based on a clinical or diagnostic examination, can often give to an employee the visual skills he needs for his job or restore the skills he has lost with increasing age.

---

[35] Extensive publications in this field are listed in bibliographies on industrial vision:

H. S. Kuhn, "Articles Bearing on Industrial Eye Problems," *Transactions of the American Academy of Ophthalmology and Otolaryngology*, 50 (1946), 175–178.

N. F. Stump, "A New Concept of Ophthalmic Service," *Bausch & Lomb Magazine*, 22 (No. 1, 1946), 4–20.

Although our primary interest in vision tests is for personnel selection and placement purposes, it should be recognized that such tests are also used for other purposes. One of these is in connection with the maintenance of employees' visual skills. If, by the use of a vision test, it is found that an employee's vision is deficient for his job, he may be referred to a physician or optometrist for professional eye care. Another use is in connection with workmen's compensation. For this purpose it is necessary to have some measure of the individual's visual skills. Thus, although we will be discussing the use of vision tests for aptitude testing purposes, their other uses should also be kept in mind.

The Nature of Visual Skills Extensive research on the visual skills of people on various jobs in industry and business has indicated certain aspects of visual performance as being most important for classifying and placing employees according to differences in visual characteristics. These aspects of visual performance are described below, along with some observations regarding tests for measuring such skills.

Keenness of vision (visual acuity) at appropriate distances (usually tested at 20 feet and 13 or 16 inches). This visual function is the ability to discriminate black and white detail, measured in terms of the minimum separable areas that can be distinguished. For industrial placement such a test should be equally valid for illiterate and literate subjects and should avoid the complications introduced by a factor involving discrimination and recognition of different shapes such as letters. The scale should make it possible to classify acuity scores adequately throughout the entire range of performance. Since acuity is modified by brightness, glare, and other external conditions, the acuity test must be given under controlled and standard conditions.

Discrimination of differences in distance (depth perception, or stereopsis). This function is an important phase of correct perception of spatial relationships. Of several cues for judging relative distances of objects, the most important for persons with normal vision, and the one that can be controlled and measured most reliably, depends on the slight difference in the position of the two eyes. The two eyes perform a geometric triangulation upon a distant object, and the distance of that object is perceived through an integration of the minute differences in appearance of the object to the two eyes. Other cues for perceiving distance in the third dimension may augment but cannot adequately substitute for this cue from two-eye functioning. Stereopsis is measurable quantitatively like most other human functions and should be measured for employee placement.

Discrimination of differences in color. Accurate color discrimination is important for efficient performance on many industrial jobs. The particular colors to be differentiated should include as many combinations as there are factors in color sensitivity. The most important combina-

tions are those that represent common colors in signal lights. In order to read such signals correctly, it is important not only that an employee should be able to differentiate between them but also that he should be able to identify and interpret correctly the meaning of each color. Any test of color vision can and should be scaled for quantitative measurement and classification of employees.

Postural characteristics of the eyes (phorias) at appropriate distances (usually 20 feet and 13 or 16 inches). Under normal seeing conditions the two eyes must move in relation to each other so that both converge symmetrically upon the object. It is this convergence that gives us a clear, single image of the object when it is viewed binocularly. Under certain testing conditions, which eliminate the necessity for such convergence of the eyes on a single point, the eyes assume a posture that may converge or diverge from that required in normal seeing at the test distance. Such postures (called "phorias" in clinical terminology) are measured in terms of angular deviation from the posture normally required for that distance. The deviation may be lateral or vertical and is measured separately in each direction. The explanation of this phenomenon has not been determined finally; nevertheless, such characteristics should be measured because they are related to performance on many industrial jobs. Such measurement must be done with adequate control and standardization of the several factors that may modify the measurement, such as the distance and focus requirements of the test.

These are not, of course, the only visual functions that are of importance in all industries, but they comprise a reasonable minimum for a program of testing applicants and employees. Tests of these functions have been adapted satisfactorily for simple and rapid use in industry, and they are the ones most widely recommended by leaders in industrial eye care for use in industry.

Examples of Vision Tests Probably the most commonly used vision test is the Snellen letter chart that is a familiar sight on the walls of physicians' offices. The test consists of several rows of block letters of decreasing size, usually placed at a distance of 20 feet from the subject. A typical Snellen chart is shown in Fig. 6.14. The test is administered by determining, separately for each eye, the smallest letters that the subject can read. The larger the letters at 20 feet that are the smallest readable, the poorer the visual acuity. The Snellen notation of acuity scores is in the form of a fraction—the smaller the fraction, the poorer the vision. In this fraction the numerator is constant and represents the distance of the test. Thus, visual acuity scored 20/20 is standard. A score of 20/40 means that the subject can read at 20 feet only a letter twice as large as standard, a letter that the "standard eye" can read at 40 feet.

The Snellen acuity designations are not intended to represent fractions of useful vision. The employee who scores 20/40 at a distance of

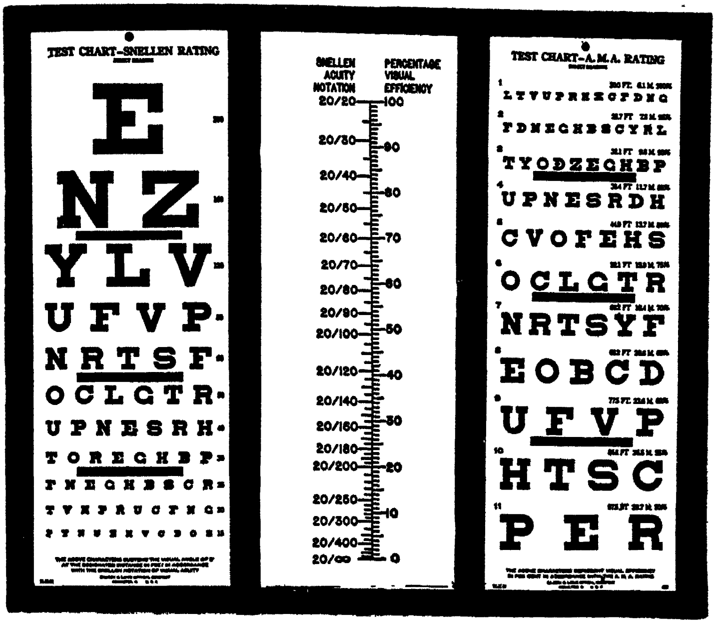


Fig. 6.14. Snellen and "A.M.A." test charts with scale for converting Snellen scores to per cent of visual efficiency.

20 feet is not necessarily handicapped 50 per cent in opportunity to earn a living. In order to simplify interpretation of visual acuity scores and to set up an equitable scale for awarding compensation in proportion to actual incapacity due to eye injury, the American Medical Association has adopted and recommended for use in industry a percentage system of acuity notation, with the distinguishing title of "Visual Efficiency." [86] Where the Snellen test measures acuity from the angles subtended by the letters, the American Medical Association notation interprets these angles in terms of percentage of visual efficiency. The difference between this percentage and 100 per cent is the percentage "loss of vision." A conversion scale for translating acuity scores into the percentage notation is shown in Fig. 6.14. Also shown is the first half of a letter chart that measures visual angles in steps directly equivalent to intervals of 5 per cent on the American Medical Association scale.

This American Medical Association percentage notation is ordinarily not carried above 100 per cent or below 20 per cent, and for compensation purposes in industry no extension of this range is necessary. Acuity

[86] Albert C. Snell and Scott Sterling, "The Percentage Evaluation of Macular Vision," *Archives of Ophthalmology,* 54 (1925), 443–461.

above standard is usually considered as only 100 per cent in computing compensation awards; and in the absence of records, 100 per cent acuity is assumed prior to the time of an eye injury. Twenty per cent acuity or less is frequently considered as complete lack of acuity or industrial blindness, and is therefore not measured or classified more precisely. These practices vary in different states, and the industrial relations officer should be familiar with the legal practices in his state regarding compensation for eye injury in industry.[37]

In the upper range of acuity scores (standard or above) adequate segregation of differences in acuity is not possible with these letter charts, since the common Snellen charts provide only two or three test levels above standard and the percentage chart provides none. Other limitations of the Snellen letter test have been pointed out, and nonliterate test charts have been developed. These charts eliminate letters and require the subject to identify a spatial pattern, such as a broken ring that may have its open area at the top, bottom, right, or left. In spite of the limitations of both the Snellen and American Medical Association charts, these tests have furnished industry with a convenient and acceptable basis for segregating and classifying cases of substandard acuity in questions of compensation for eye injury. Equivalent information may be obtained, of course, from more elaborate vision tests used in the plant for other purposes.

Because such tests as the Snellen and the American Medical Association do not adequately serve selection and placement purposes, various other vision-testing devices have been developed. Such devices measure various visual skills in addition to visual acuity, and also provide for the testing of people at the optical equivalent of both near and far distances. Four such instruments are the Ortho-Rater, the Sight-Screener, the Telebinocular, and the Vision Tester.[38]

The Bausch & Lomb Visual Classification and Placement Tests for Industry[39] are the first battery of vision tests to be constructed on the basis of specifications derived from extensive investigations among industrial employees in industrial situations. These tests cover the visual functions described previously and, for maximum speed and convenience in testing, are incorporated in a single instrument, shown in Fig. 6.15. This instrument, called the Ortho-Rater, is a precision stereoscope of relatively long focal length that permits adequate and separate control of test stimuli for each eye. Tests are given at optical equivalents of 26

---

[37] See A. C. Snell, *Medicolegal Ophthalmology* (St. Louis: The C. V. Mosbey Company, 1940).

[38] The Ortho-Rater is manufactured by Bausch & Lomb Optical Company, Rochester, New York; the Sight-Screener by American Optical Company, Southbridge, Massachusetts; the Telebinocular by Keystone View Company, Meadville, Pennsylvania; and the Vision Tester by the Titmus Optical Co., Petersburg, Virginia

[39] The Bausch & Lomb Optical Company, Rochester, New York.

Fig. 6.15. The Ortho-Rater in use in a hosiery mill. (Courtesy Haynes Hosiery Mills Company, Winston-Salem, N.C.)

feet and 13 inches. Stereoscopic methods of vision testing have been used since the late nineteenth century and were early described by Wells.[40] The early statistical data upon which these tests are based have been published separately in technical and professional journals.[41]

The Ortho-Rater includes the following 12 tests: far-distance tests of vertical phoria, lateral phoria, acuity both eyes, acuity left eye, acuity right eye, depth perception, and color discrimination; near-distance tests of vertical phoria, lateral phoria, acuity both eyes, acuity left eye, and acuity right eye.

Examples of Vision Tests in Use Extensive research carried out over a period of years at the Occupational Research Center of Purdue Univer-

[40] David Wells, *The Stereoscope in Ophthalmology* (2nd ed.; Globe Optical Company, 1918).

[41] W. J. Giese, "The Interrelationship of Visual Acuity at Different Distances," *Journal of Applied Psychology,* **30** (1946), 91–106.

F. W. Jobe, "Instrumentation for the Bausch & Lomb Industrial Vision Service," *Bausch & Lomb Magazine,* **20** (1944), No. 2, 3–5, 19–21

Joseph Tiffin and H. S. Kuhn, "Color Discrimination in Industry," *Archives of Ophthalmology,* **28** (1942), 851–859.

Joseph Tiffin and S. E. Wirt, "Near *vs.* Distance Visual Acuity in Relation to Success on Close Industrial Jobs," *Supplement to Transactions of the American Academy of Ophthalmology and Otolaryngology,* June 1944, 9–16.

S. E. Wirt, "The Validity of Lateral Phoria Measurements in the Ortho-Rater," *Journal of Applied Psychology,* **27** (1943), 217–232.

sity, and elsewhere, has demonstrated conclusively that jobs differ in the visual demands they make upon the worker. These variations are both qualitative and quantitative. Some jobs, for example, require the ability to see at a distance, such as in operating a crane or driving a truck; and others require the ability to see close at hand, such as in watchmaking or fine assembly work. These are qualitative differences, since they are predicated on different visual skills. In addition, jobs vary in the amount of a given skill that is required; some jobs, for example, require greater color discrimination than others.

It should be added here that the correlations among visual skills are not particularly high. It is therefore not possible to predict one visual skill from a knowledge of a person's score on a different type of vision test. Being able to read the chart across the doctor's room gives no assurance of being able to read the fine print on the insurance policy.

In the industrial vision research conducted at the Occupational Research Center, it was found that vision tests could be more effectively used if visual standards for jobs were set up in the form of "visual profiles," each profile specifying the minimum acceptable scores on the various Ortho-Rater tests. If a person had scores above the specified minimums, he "passed" the profile, and if he had one or more scores below any specified minimum, he "failed" the profile. Profiles for individual jobs were developed, using various criteria such as quantity of work, quality of work, ratings, and accidents.

Visual job families. Over a period of years data were collected on vision test scores and criteria of job success for individuals on several thousand jobs. A careful analysis of the relationships between vision test scores and job success as they have been revealed in these many different job situations has shown that there are in industry certain groups of jobs that are similar to each other in terms of visual requirements. Within each group the visual requirements of the jobs are similar, but the requirements of each group vary from those of the next group. Thus we have a series of job groupings each representing a different pattern of visual requirements although the jobs within each group are essentially similar in visual requirements. This has led to the concept of "visual job families." A visual job family is composed of a group of jobs whose visual requirements are similar. The requirements between families, however, differ.

Six such visual job families have been identified. Research has shown that the vast majority of industrial jobs will fall into one or another of these six groups in terms of the visual demands the job makes upon the worker. The visual requirements in terms of Ortho-Rater test scores of each of the six job families are shown in Fig. 6.16. Descriptions of these families are given below:

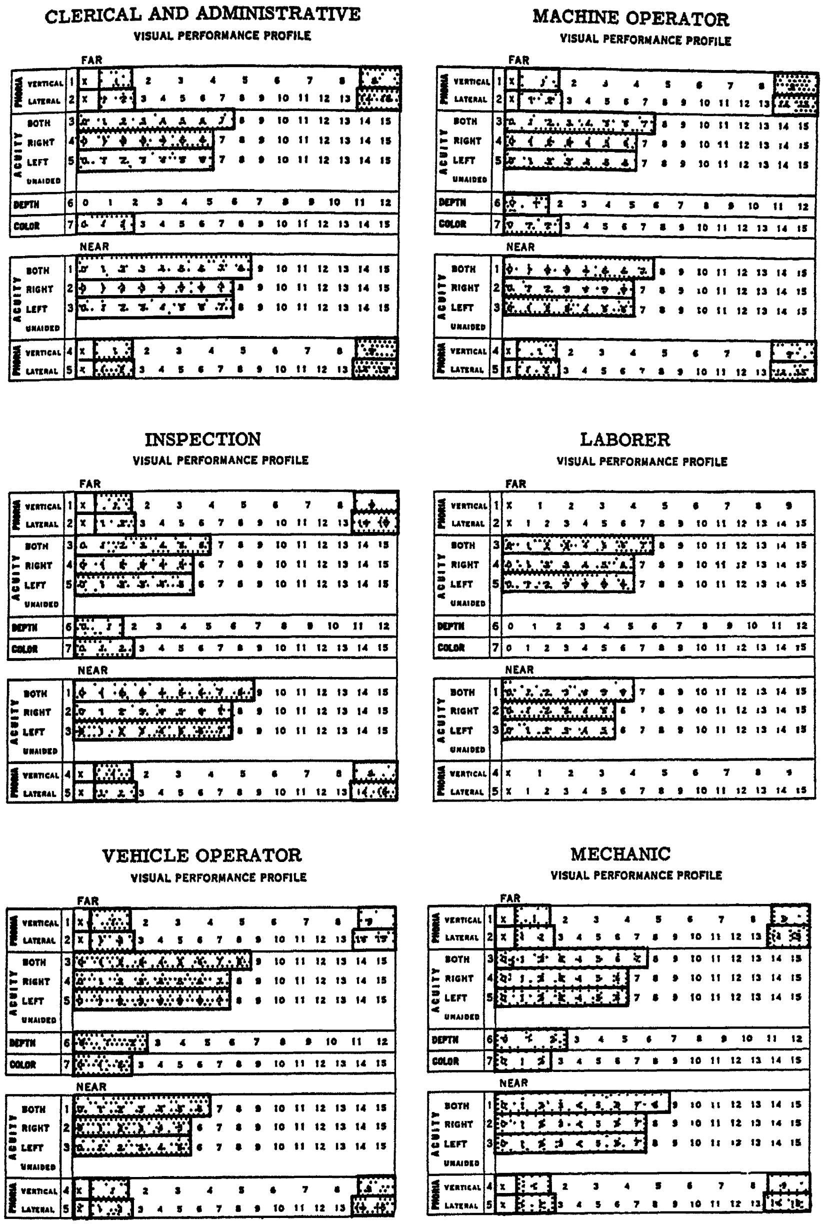


Fig. 6.16. Visual skill test requirements for the most common job groups. An employee "passes" the visual requirements if all of his Ortho-Rater scores fall in unshaded areas. He "fails" if one or more scores fall in a shaded area.

Visual Job Family I. Clerical and administrative. This standard covers those jobs primarily concerned with paper work. All types of clerical jobs and those administrative occupations that are of the desk-work type are included.

Visual Job Family II. Inspection and close work. This standard covers jobs involved in the inspection of small parts for surface defects. Also involved are jobs of the machine-operating type in which the work is done at close range (such as sewing-machine operator). Assembly jobs involving very small parts (such as watches, radio tubes, and so forth) also fall in this category.

Visual Job Family III. Vehicle operator. This standard covers jobs requiring the operation of moving vehicles (truck driver, crane operator, high-lift operator, and so forth).

Visual Job Family IV. Machine operator. This standard covers those jobs involving the operation of machines in which the operating parts of the machine are within arm's length (such as lathes, drill presses, spinning machines, and so forth).

Visual Job Family V. Laborer. This standard involves jobs of the relatively unskilled type (porters, janitors, guards, hand truckers, and so forth).

Visual Job Family VI. Mechanic and skilled tradesman. This standard involves jobs of the mechanical type (such as radio mechanic, Diesel mechanic, machine fixer, and so forth). Also included are the skilled trades (such as carpenter, plumber, millwright, electrician, and so forth).

In order to investigate the relationship between visual skill test score requirements as represented by these six visual standards and success on industrial jobs the method of cross-validation was used. A random sample of 43 individual job studies was selected. Each unit of this sample represented a group of data involving visual test scores and measures of job success. From the job-description materials supplied by the companies, the jobs were assigned to one of the six visual job families. The workers were then divided into those whose visual skills were adequate for the job and those whose visual skills were not adequate for the job, using the visual skills profile appropriate for this job family. For each of these visual groups the percentage of workers who were in the high-criterion group on the basis of job success was computed. Thus each unit in this sample represented an individual study of the relationship between vision and job success in an individual job situation in a specific plant.

Table 6.3 presents the results of these studies grouped by visual job families. This table shows the number of jobs classified into each of the six job families, the number of employees in each of these groups, and the per cent of "high-criterion" employees in each group among those whose vision is adequate for their jobs and the corresponding per cent of

Table 6.3

RELATIONSHIP BETWEEN VISUAL SKILL PROFILES AND SUCCESS ON INDUSTRIAL JOBS IN EACH OF SIX VISUAL JOB FAMILIES

Visual Job Family	*No. of Jobs*	*No. of Employees in Job Family Group*	*Per cent of "high-criterion" employees on the jobs among those who have: Adequate Vision*	*Inadequate Vision*
I	5	170	71%	37%
II	9	783	62%	50%
III	5	391	59%	45%
IV	10	497	63%	45%
V	5	234	67%	34%
VI	9	345	69%	57%
All Job Families	43	2,420	65%	46%

"high-criterion" employees among those whose vision is inadequate for their jobs.

The obvious conclusion from the results shown in Table 6.3 is that there is consistently a higher per cent of high-criterion employees among those whose vision is adequate for their jobs than among those whose vision is lacking in this respect. It shou~~ld be kept in mind~~ that satisfactory vision for one visual job family does not at all guarantee adequate vision for another type of job.

It should be emphasized again that the 43 jobs included in this study had not been used in any way in establishing the visual job family profiles. The profiles used (shown in Fig. 6.16) had already been established before the data on the 43 jobs were obtained. These jobs therefore represent strictly "hold-out" groups and the present investigation was a cross-validation of the previously established job family visual skill profiles.

The degree of discrimination among the 43 jobs varied a great deal, as would be expected, such variations being in part due to the relative importance of visual skills for the jobs in question. In general, however, the findings present very persuasive evidence that there is a definite relationship between vision test scores and success on many industrial jobs.

TEST OF SPECIAL APTITUDES

Aside from the types of aptitude tests discussed above, there are many others that deal with other special aptitudes or attributes. Such special tests sometimes have unique applicability for particular purposes. Among such tests are those of clerical aptitude, reading speed and comprehension, vocabulary, and perceptual speed. In addition, there are certain

series of tests, each of which includes sub-tests of various aptitudes or other abilities. Some such tests are listed in Appendix D. For illustrative purposes here, one type of special aptitude test will be mentioned briefly, namely, reading tests.

Reading tests. While reading ability (both speed and comprehension) is markedly related to intelligence, it is in order to discuss tests in this area under our category of special abilities because some people who score very high on mental tests are not outstanding in their ability to read. This is undoubtedly due to the fact that their reading training was inadequate, but the fact remains that if their reading skills are not at a high level, they will be handicapped in undertaking certain kinds of work no matter how intelligent they are. For this reason, reading tests have found a real use in industry. Both of the two aspects of reading skills —speed and comprehension—are important. With a mass of printed material flowing across every supervisor's desk, the man who does not read rapidly soon loses touch with his job. He needs to read rapidly and with satisfactory comprehension. There is not a "speed of reading" test available that was constructed specifically for use with industrial employees or supervisors, but the Purdue Reading Test for Industrial Supervisors [42] was constructed specifically to measure reading comprehension among supervisors in the industrial setting. It consists of 13 paragraphs, each followed by several objective questions dealing with the material covered in the paragraphs. A generous time limit of 25 minutes enables most men to finish the test in the time allowed. The reliability of the test is .91, and the scores correlate .81 with reading test scores obtained with the Nelson Silent Reading Test for Grades 3 to 9.

The results of one investigation of the validity of the Purdue Reading Test are shown in Fig. 6.17. As this figure shows, there are far more superior quality control supervisors among men who score high on the test than there are among the low-scoring men.

VALIDITY OF APTITUDE TESTS FOR VARIOUS OCCUPATIONS

In the preceding sections of this chapter various studies have been mentioned that deal with the validity of aptitude tests for specific jobs. It may now be in order to present an over-view of the validity of different kinds of aptitude tests for the selection of people for various occupational groups. Two studies by Ghiselli and Brown [43] summarize

---

[42] Joseph Tiffin and Roy Dunlap, "Purdue Reading Test for Industrial Supervisors," University Book Store, West Lafayette, Ind.

[43] E. E Ghiselli and C. W. Brown, "Validity of Aptitude Tests for Predicting Trainability of Workers," *Personnel Psychology*, 4 (1951), 243–260; C. W. Brown and E. E Ghiselli, "Some Generalizations Concerning the Validity of Aptitude Tests," *Personnel Psychology*, 6 (1953), 139–150.

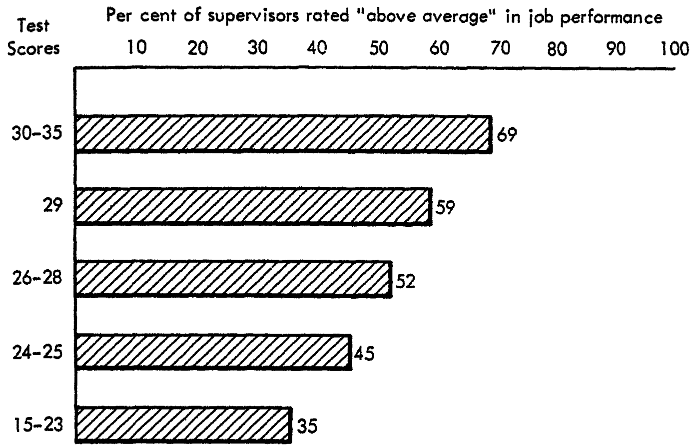


Fig. 6.17. Per cent of Quality Control Supervisors rated "above average" in job performance in relation to scores made on the Purdue Reading Test for Industrial Supervisors.

the results of many published and unpublished test-validation studies in a manner that provides us with such an over-view. The first summary deals with studies relating to the prediction of performance in occupational training courses and of apprentices in training. The second summary deals with studies relating to the prediction of job proficiency. The second summary is particularly pertinent; it will be described briefly.

For each test-validation study that was included, information was available on (1) the job, (2) the test or tests used, (3) the criterion, and (4) the correlation between test scores and criterion measures (test validity). All of the specific tests used were originally classified into 16 different categories. Subsequently, on the basis of an analysis of the relationships among the tests, the tests were grouped into the following three "clusters":

A. *Intellectual ability* (including tests of intelligence, memory, substitution, arithmetic, number comparison, name comparison, and cancellation).
B. *Spatial ability* (including tests of tracing, location, pursuit, and spatial relations).
C. *Motor ability* (including tests of tapping, dotting, and finger, hand, and arm dexterity).

The specific jobs were then classified into 21 job groups, as shown in Fig. 6.18.

Then, by a method of analysis that need not be described here, the "relative validity" of each of the three types of tests was determined for

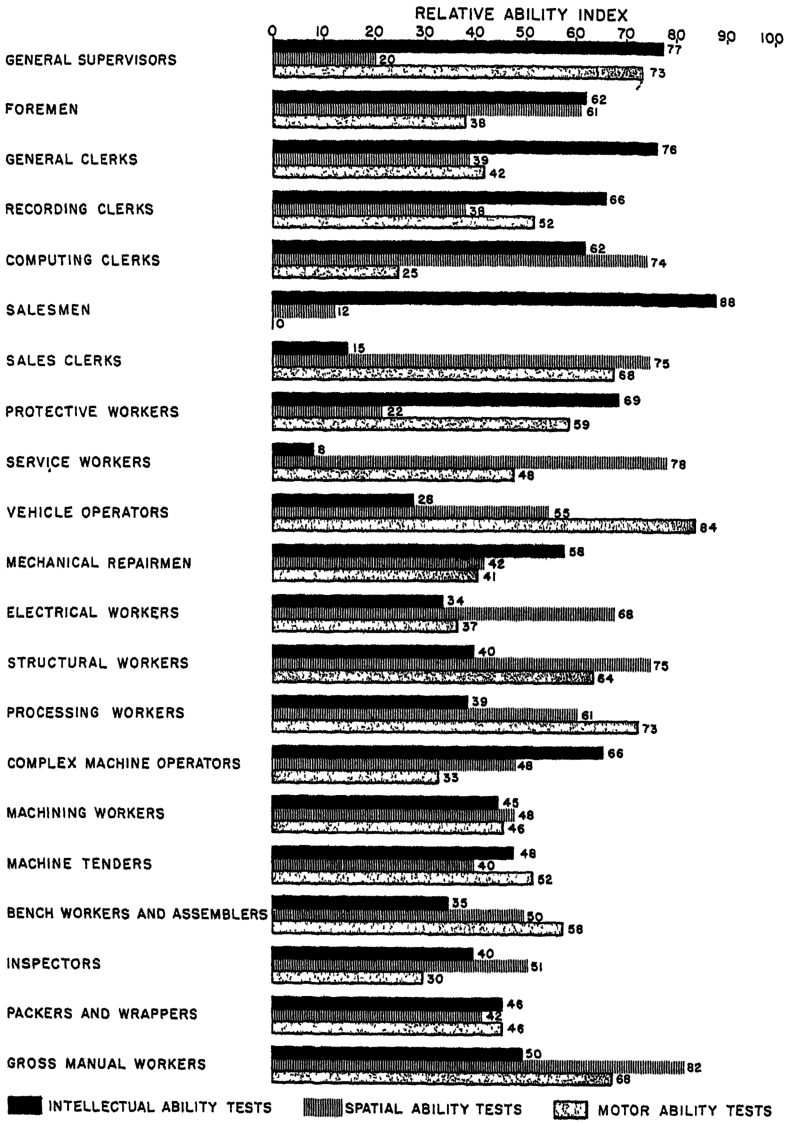


Fig. 6.18. "Relative validity" index of three types of tests for various classes of jobs. (Adapted from Brown and Ghiselli, *op. cit.*)

each of 21 job categories. These are expressed in terms of an arbitrary scale,[44] which *cannot* be interpreted as a correlation. In other words, this figure simply shows an estimate of the *relative* validities of three types of tests for the various job classes.

It will be seen that intellectual ability had the highest relative validity for general supervisors, foremen, the clerical groups, salesmen, protective-service workers, and complex-machine operators. This ability has little relationship to success on such jobs as salesclerks, service workers, and vehicle operators.

The spatial ability tests showed the highest relative validity indexes for structural workers (which would be expected), and for some jobs for which such indexes would not be expected, namely, computing and sales clerks, service workers, and gross manual workers. The validity indexes of spatial ability were lowest for general supervisors, salesmen, protective workers, and general and recording clerks. Most of the mechanical jobs have intermediate validity.

The motor ability tests also revealed some unexpected results. The high values are logical for such job groups as vehicle operators, processing workers, and possibly gross manual workers, but are not as easily explained for protective workers and salesclerks. For most of the industrial jobs, the validity indexes were of intermediate values.

Some of the unexpectedly low validity indexes (such as for the spatial ability test cluster for mechanical jobs) may be explained in part by "negative selection" within the groups covered by the study. The voluntary or involuntary elimination of some workers on a job (particularly the poorer workers) frequently tends to reduce the range of test scores, and also to reduce, spuriously, the correlation of the test with the criterion. This effect, in turn, may have caused the *relative* validities for other job groups to be somewhat higher than they really should be.

Although the results summarized need to be reviewed with certain reservations (such as the one mentioned above), it nonetheless helps to summarize a great deal of the work that has been carried out in validating aptitude tests. In particular, it can give cues regarding the relative importance of various types of aptitude tests for the selection of people for each of several job groups.

44 The scale is a standard score scale with a mean of 50 and a standard deviation of 20.

7

Tests of Personality and Interest

Employment managers universally recognize the importance of personality traits in employees whom they hire. Indeed, one of their reasons for sometimes being hesitant in adopting psychological tests is that they often think of tests only in terms of intelligence or dexterity, and these tests do not, of course, take into account the more general personality traits of the applicant. An applicant might be very high in mental ability or in manipulative dexterity and yet have a personality that would not only unfit him for the job for which he is applying but would also make him a definitely undesirable individual to employ in any capacity.

The foregoing sections, which have dealt with aptitude and achievement tests without regard to the applicant's general personality, do not imply that the more general personality traits are unimportant. Psychologists are the first to recognize the importance of personality traits in helping an employee adapt himself to any job or to any organization. Psychological tests have emphasized specific aptitudes because psychologists recognize the importance of job aptitude, as such, aside from personality characteristics, and because, up to the present time, it has proved possible to develop adequate tests of such aptitudes as finger dexterity or intelligence to a greater extent than it has been possible to develop adequate tests in the complicated field of personality.

Within recent years, however, both employment managers and consulting psychologists have increasingly demanded some reasonably satisfactory and accurate method of determining certain personality traits of an applicant at the time he applies for employment. It is now recognized not only that an applicant who does not have the aptitude to learn the job will fail, no matter how desirable his personality traits may be, but also that if he does have the aptitude for the job he will probably still fail if his personality characteristics make it difficult for him to fit into the organization and to work cooperatively with other persons. Although this conclusion is sound common sense, we do not need to rely entirely upon subjective judgment to reach this conclusion. Various investigations have shown that personality and adjustment factors are responsible for the ineffectiveness of job performance of many people, rather than lack of relevant aptitudes or technical competence. Thus, it is desirable to be able to identify, at the time of employment, those individuals who have the personality traits that would be conducive to satisfactory job performance.

Similarly, there is increasing recognition of the importance of interests of employees as factors that are associated with job performance. Comparing persons on similar jobs, there is reason to believe that those who do *not* like their jobs are not as good in job performance, and do not stay as long, as those who *do* like their jobs. Moreover, the individuals who are doing work in which they have little or no interest do not have the opportunity to achieve the level of job satisfaction that is experienced by persons who are doing what they like to do.

It has been with the hope of measuring personality traits and interests that psychologists have developed various personality and interest tests. In the development and use of such tests, however, difficulties have been encountered which at present place serious limitations on their use for employment purposes. In particular, these limitations stem from the possibility of "faking" on the test, that is, the possibility of an applicant giving responses to test items that are not descriptive of himself. More will be said about this later, but first it would be desirable to describe briefly some of the personality and interest tests that are available.

TESTS OF PERSONALITY

Most of the personality tests being used in industry are paper-and-pencil tests, with multiple-choice items. Actually, they are not "tests" at all, in the sense that there are correct and incorrect answers to the questions. They are really "questionnaires," in which the applicant or employee is asked to tell how he feels about or would react in certain specified situations. The items typically consist of statements, or questions,

relating to behavior, attitudes, or beliefs. A typical item is given below:

1. I feel uncomfortable with other people.
   a) Yes
   b) Don't know
   c) No

The responses typically are scored to give a score on each of several personality or temperament "traits," such as "dominance," or "sociability." The responses that are "scored" for a particular trait or personality attribute usually have been identified by statistical item analysis or on rational grounds as being responses that are characteristic of those individuals who have that trait or attribute. The scores on the various traits or attributes usually are treated as individual test scores in the same manner as scores on other types of tests.

In a few companies "projective" tests have been used. These usually are individual (rather than group) tests. In a projective-type test the subject is presented with each of several items such as ink blots or pictures, and he is asked to tell what he sees in the object in question. A trained test administrator interprets the individual responses and the patterns of responses to the various items, and makes an assessment of the individual. Usually this is in qualitative rather than quantitative terms.

Specific Personality Tests Among the personality tests that have been used in industry are the following:

The Bernreuter Personality Inventory
The Humm-Wadsworth Temperament Scale
The Guilford Series of Personality Tests
The Edwards Personal Preference Schedule
The Minnesota Multiphasic Personality Inventory
The Thurstone Temperament Schedule
The Gordon Personal Profile
The Gordon Personal Inventory
The Rorschach Projective Test

These are listed in Appendix D. In order to indicate further the nature of personality tests, however, one such test will be described briefly.

The Minnesota Multiphasic Personality Inventory.[1] This inventory

[1] Distributed by the Psychological Corporation. The research on which the scales are based is reported in the following articles:

S. R. Hathaway and J. C. McKinley, "A Multiphasic Personality Schedule (Minnesota): I. Construction of the Schedule," *Journal of Psychology,* 10 (1940), 249–254; "II. A Differential Study of Hypochondriasis," *Journal of Psychology,* 10 (1940), 255–268; "III. The Measurement of Symptomatic Depression," *Journal of Psychology,* 14 (1942), 73–84; "IV. Psychasthenia," *Journal of Applied Psychology,* 26 (1942), 614–624.

J. C. McKinley and S. R. Hathaway, "The Minnesota Multiphasic Personality Inventory: V. Hysteria, Hypomania, and Psychopathic Deviate," *Journal of Applied Psychology,* 28 (1944), 153–174.

consists of 550 statements that the person being tested sorts into three piles—those he regards as *true,* those he regards as *false,* and those on which he cannot make a true or false judgment. After the cards have been sorted, the scoring procedure yields eight different personality component scores, viz.:

> *Hypochondriasis Scale.* Measures undue worry about one's health.
>
> *Depression Scale.* Measures feeling of unworthiness, uselessness, and undue worry.
>
> *Hysteria Scale.* Measures the degree to which subject is like patients who have general systemic complaints and/or specific complaints such as paralyses, contractures, and gastric, intestinal, and/or cardiac symptoms.
>
> *Psychopathic Deviate Scale.* Measures similarity of subject to patients whose main difficulty lies in their absence of deep emotional response, inability to profit from experience, and disregard of social mores.
>
> *Interest Scale.* Measures tendency toward masculinity or femininity of one's interest pattern.
>
> *Paranoia Scale.* Measures stubborn adherence to fixed ideas, suspicion, and contempt for the opinion of others
>
> *Psychasthenia Scale.* Measures similarity of the subject to psychiatric patients who are troubled with phobias (unreasonable fears) or compulsions.
>
> *Schizophrenia Scale.* Measures similarity of subject to psychiatric patients who are characterized by bizarre and unusual thoughts and behavior.

In addition to the eight personality components measured, the Minnesota Multiphasic Personality Inventory also provides for certain checks on the accuracy and honesty of the subject's responses. A further expansion of the scoring possibilities of the scale has been developed by Drake[2] in the form of a key to obtain scores on social introversion-extroversion. Simplified scoring of the scale for group administration, to be used when electric, punched-card tabulating equipment is available, has been described by Burton and Bright.[3]

An indication of the validity of the Minnesota Inventory comes from the work of Schmidt,[4] who reports that the Inventory "distinguished . . . much statistical significance between normal soldiers and those diagnosed as constitutional psychopaths; mild or severe neurosis; and psychosis."

INTEREST TESTS

Interest tests typically require the person being tested to indicate the strength of his interests in such things as various jobs, hobbies, recreation, or leisure-time activities. This is sometimes done by presenting

---

[2] L. E. Drake, "A Social I. E. Scale for the Minnesota Personality Inventory," *Journal of Applied Psychology,* **30** (1946), 51–54.

[3] A. Burton and C. J. Bright, "Minnesota Multiphasic Personality Inventory for Group Administration," *Journal of Consulting Psychology,* **10** (1946), 99–103.

[4] H. O. Schmidt, "Test Profiles as a Diagnostic Aid: The Minnesota Multiphasic Inventory," *Journal of Applied Psychology,* **29** (1945), 115–131.

groups of, say, three activities, and by asking the individual which he likes most, and which he likes least. In other cases, he simply indicates, for each statement, the extent to which he likes or dislikes the activity in question.

Specific Interest Tests While there are a number of standardized interest tests available, the two that are in most common use are the Strong Vocational Interest Blank and the Kuder Preference Record.

Strong Vocational Interest Blank.[5] This test has been developed primarily for use in vocational guidance counseling. It determines whether the subject's pattern of interests agrees with the interest pattern of men in each of a large number of professions and occupations. Other things being equal, a person choosing a profession or occupation is more likely to be happy and successful if his basic interests are similar to the interests of men actually in this particular field.

Different forms of this test and different scoring keys are available, so that it may be used with men or women of student or adult age.

The Strong Vocational Interest Blank measures as many patterns of interest as vocational patterns for which it is scored. This procedure is advisable in vocational guidance work, and the test is admirably suited for that purpose.

Kuder Preference Record. There are four forms of the Kuder Preference Record, one of which will be described. Form C (Vocational) measures ten broad areas of interests, as follows:

Outdoor
Mechanical
Computational
Scientific
Persuasive
Artistic
Literary
Musical
Social Service
Clerical

Separate scores are readily obtained for each of the listed interest groups. This type of scoring makes the test particularly easy to try out on a "test the test" basis (see p. 119) because only the ten basic scores will be considered, and these yield an over-all pattern of interest. The manual with this test, however, provides interpretive data for evaluating various interest patterns in terms of their agreement with those of men in various occupations.

Just as in the case of personality tests of the type previously discussed,

[5] This test is distributed by the Psychological Corporation.

the scores on an interest test, such as the Kuder Preference Record, may be distorted by an applicant or employee who attempts to "fake" his answers. Kuder[6] has described two methods of scoring the Preference Record to differentiate between blanks answered sincerely and blanks answered with the purpose of making a good impression. These scores were found to identify faked *vs.* sincerely answered blanks with from 87 to 94 per cent accuracy.

LIMITATIONS OF PERSONALITY AND INTEREST TESTS

As implied above, personality and interest tests have serious limitations for employment purposes, the most important one being the possibility of faking on the part of applicants.

Faking of Personality and Interest Tests If a person is taking a personality or interest test for the purpose of receiving vocational guidance, he will usually be motivated to give relatively truthful answers to the test questions, since it is to his interest to find out all he can about himself in order to make a sound vocational choice. If he is applying for a job, however, his motivation to get the job may induce him to give responses that he *thinks* will make him out to be the kind of person the company is looking for.

Evidence of the ability of people to fake such tests is available from various studies. One such study, for example, was carried out with the Humm-Wadsworth Temperament Scale with 65 college students.[7] Each student was given the scale twice: first, in a "clinical" situation, with instructions to be as frank as possible, and second, in an "employment" situation in which each student was to assume that he was in an employment office after a job and had been asked to take the test as a part of the employment procedure. Table 7.1 shows the mean scores for the seven components obtained under these two conditions.

It is apparent from Table 7.1 that the employment situation, when compared with the clinical, shows a higher average value for the normal component and lower values for all except the epileptoid of the remaining six components. In other words, the students were able, by assuming an attitude of "applying for a job," to change their test profiles toward more of the normal and less of the undesirable traits. All differences shown in Table 7.1 are significant from a statistical viewpoint, and only those scores were used in the computation that fell within the no-count limits, for both the clinical and employment situations, within which the manual of instructions states that the scale should be "accepted as probably valid."

Even with the shift in means from one situation to the other, it would

[6] G. F. Kuder, "Identifying the Faker," *Personnel Psychology*, 3 (1950), 155–167.

[7] This study was conducted by W. J. Giese and F. C. Christy at Purdue University.

Table 7.1

MEAN SCORES ON HUMM-WADSWORTH TEMPERAMENT SCALE OBTAINED IN A CLINICAL AND AN EMPLOYMENT SITUATION BY 65 COLLEGE STUDENTS *

Component	*Clinical Mean*	*Employment Mean*	*Shift from Clinical to Employment*
Normal	981	1023	+ 42
Hysteroid	1023	980	− 63
Manic	1035	937	− 98
Depressive	1061	913	−148
Autistic	1024	938	− 86
Paranoid	970	955	− 15
Epileptoid	983	1002	+ 19

* *All scores were computed by the log method, with correction for no-count, as described in the manual of directions.*

still be possible to infer one's score in one situation from a knowledge of his score in the other situation if scores in the two situations were highly correlated. The correlations were computed and are shown in Table 7.2.

Table 7.2

CORRELATIONS BETWEEN CLINICAL AND EMPLOYMENT SITUATIONS ON THE HUMM-WADSWORTH TEMPERAMENT SCALE FOR 65 COLLEGE STUDENTS

Component	*Correlation*
Normal	−.03
Hysteroid	+.42
Manic	+.09
Depressive	−.10
Autistic	+.11
Paranoid	+.61
Epileptoid	+.23

The only correlation in Table 7.2 that is large enough to be significant is the one for the paranoid component.

Without going into any details, it might be said that other somewhat corresponding studies with other tests have given similar results. For example, Longstaff [8] found that both the Strong Vocational Interest Blank and the Kuder Preference Record could be faked to a considerable degree, although some interest categories were more fakable than others.

[8] H. P. Longstaff, "Fakability of the Strong Interest Blank and the Kuder Preference Record," *Journal of Applied Psychology*, 32 (1948), 360–369.

In another study, Longstaff and Jurgensen[9] found that scores on the Jurgensen Classification Inventory could be changed significantly, depending on the mental "set" of the individuals at the time as affected by the instructions given to the students taking the test.

The fact that people can fake personality and interest inventories implies serious limitations to the potential users of such tests for employment purposes. There is, however, the possibility that tests can be developed that are not as fakable as some of the current tests. The most promising development along these lines is the use of forced-choice tests.

Fakability of Forced-Choice Tests The Gordon Personal Profile, the Gordon Personal Inventory, and the Edwards Personal Preference Schedule are based on the so-called forced-choice technique. By this method, the items that are put together within a group are matched in terms of their favorability. If there are two items in a group, one of the items discriminates in terms of the attribute that the test is supposed to measure, whereas the other item does not. (In some variants of the forced-choice technique there are more than two items in a group, in which case more than one item may be "discriminating." In other variants there are both favorable and unfavorable items; in such cases the favorable ones are equally favorable, and the unfavorable ones are equally unfavorable, and the respondent selects the response that is "most" applicable to him and the one that is "least" applicable.) Since the respondent cannot choose one response because it "looks" better (since they are equal on favorability), he presumably would tend to choose the response that is most like him. If this is a "discriminating" item, it would add to his score on the attribute in question.

Since the forced-choice technique is designed specifically to minimize faking, it is reasonable to wonder whether, in fact, the method does accomplish this objective. Some evidence comes from two studies with the Gordon Personal Profile, one by Rusmore[10] and the other by Gordon and Stapleton.[11] In the first of these studies the same group of 81 college students took the test under a simulated "industrial" situation and under a simulated "guidance" situation. In the second study 121 high school students first took the test in connection with the regular guidance program, and three months later took the test when actually applying for ouside summer employment. The results of these two studies are summarized in Table 7.3. This shows the means of the scores on the various

---

[9] H. P. Longstaff and C. E Jurgensen, "Fakability of the Jurgensen Classification Inventory," *Journal of Applied Psychology,* 37 (1953), 86–99.

[10] J. T. Rusmore, "Fakability of the Gordon Personal Profile," *Journal of Applied Psychology,* 40 (1956), 175–177.

[11] L. V. Gordon and E. S. Stapleton, "Fakability of a Forced-Choice Personality Test Under Realistic High School Employment Conditions," *Journal of Applied Psychology,* 40 (1956), 258–262.

test components for the two administrations for each group. In the case of the college students it can be seen that the simulated "industrial" situation resulted in a significant difference on the *Responsibility* and *Total* scores, although these differences were only about 9 and 8 per cent respectively. In the case of the high school students significant differences occurred in these two scores as well as in the *Emotional Stability* scale. The difference in the case of the *Responsibility* scale is rather substantial; the differences in the *Emotional Stability* and *Total* scores are more moderate (although statistically significant).

Table 7.3

MEAN SCORES ON THE GORDON PERSONAL PROFILE TAKEN TWICE UNDER SPECIFIED TEST CONDITIONS

	Group and Condition			
	College Students (81)		*H. S. Students (121)*	
Test Component	*Simulated "Guidance"*	*Simulated "Industrial"*	*Guidance*	*Employment*
Ascendancy	3.7	4.4	2.9	3.2
Responsibility	7.4	8.9 †	4.3	6.8 †
Emotional Stability	7.3	8.0	5.3	6.8 †
Sociability	4.4	4.8	4.8	4.5
Total	23.0	26.1 *	17.2	21.3 †

* *Difference between the "Guidance" and "Industrial" or "Employment" mean scores significant at the 5% level of confidence.*
† *Difference between the two means significant at the 1% level of confidence.*
Source *Adapted from articles by Rusmore and by Gordon and Stapleton,* op. cit

Another study, with the Edwards Personal Preference Schedule,[12] indicates that this test may be subject to at least some "faking." The test was administered to 362 male sales applicants of a large midwestern industrial concern. The results obtained with this group were compared with the college male group on which Edwards [13] based his norms. Large and significant differences were found between the two groups on nearly all of the 15 scales covered by the test. The sales applicant group was shown by the test to be more outgoing, dominant, ambitious, and persistent. Since these traits are traditionally considered to be characteristic of salesmen, it seems quite possible that these sales applicants were able to "beat" the test by showing themselves as having a "sales personality."

A study reported by Dicken [14] also shows that the Edwards Personal

[12] W. K. Kirchner, M. D. Dunnette, and Nancy Mousky, "Use of the Edwards Personal Preference Schedule in the Selection of Salesmen," *Personnel Psychology*, 13 (1960), 421–424.

[13] See Appendix D.

[14] C. F. Dicken, "Simulated Patterns on the Edwards Personal Preference Schedule," *Journal of Applied Psychology*, 43 (1959), 372–378.

Preference Schedule may be "faked." The test was first administered to college students with standard instructions. It was then readministered to sub-groups of the same students after the students in each sub-group had been instructed to make the best impression they could on a trait that was described to them. Large and reliable differences were found in three of the four scales when they were administered with these instructions. And just trying to make a good impression when the students took the test the second time also resulted in substantial score changes.

In considering studies of faking of forced-choice personality tests, it should be kept in mind that certain of these involved high school or college students (and in some cases in simulated situations). It is, therefore, a bit risky to assume that the faking in actual industrial situations would be of comparable type and degree. Such studies do indicate, however, that forced-choice personality tests are not free from the possibility of faking, although the extent to which they actually are faked is generally unknown. Granting that such tests have not completely resolved this problem, it is still probable that they are less amenable to faking than are conventional personality tests. It should also be suggested that with further development and refinement of tests of this type, they might be less vulnerable to faking than are those now available.

THE USE OF PERSONALITY AND INTEREST TESTS

There is a great temptation on the part of management to use personality and interest tests without taking necessary precautions in their interpretation. The following type of "reasoning" represents a fairly common mistake in adopting such tests for employment purposes:

1. One personality component measured by a given test is called *neurotic tendency*. This test *must* measure *neurotic tendency* because that title is used by the authors of the test. (The reader may substitute any presumably undesirable personality trait he wishes for the phrase *neurotic tendency*. The result of the "reasoning" will be the same.)
2. We do not want in our plant people who have a high degree of *neurotic tendency*.
3. Therefore we do not want in our plant anyone who scores beyond a certain point on this test.

The same "reasoning" can be applied to any other "undesirable" trait or interest, or can be applied in reverse to any "desirable" trait or interest.

Anyone familiar with the statistical procedures that should be followed in developing and standardizing any test for use in a given situation will recognize a fallacy in every one of the three steps given above. First, a test does not necessarily measure a given trait merely because the name of that trait is used in the title or description of the test. Second,

one cannot assume that persons with *any* given trait, *as revealed by a score on a test,* are by virtue of that fact alone necessarily undesirable employees. It follows that the third statement, based upon the accepted truth of the first and second, is hardly justified without more specific evidence on the applicability of the test to the situation in question.

The fact that many personnel managers too often place far more credence in personality tests than is justified was shown in a study reported by Stagner.[15] Stagner administered a standard personality test to 58 personnel directors who were attending a personnel conference. The men were told that the tests would be scored and each man furnished with his personality analysis. However, instead of making a separate analysis for each man, they were all given the very same "analysis," with each man's name written at the top of the sheet. The analysis consisted of a series of flattering statements taken from a variety of sources. Fifty per cent of the personnel directors stated that the "analysis" was *amazingly accurate;* 40 per cent said that it was *rather good;* and only 10 per cent said it was *about half and half.* No one said it was *more wrong than right* or *almost entirely wrong.* Stagner points out that a personnel manager should avoid being duped by a flattering report on his own fine qualities into purchasing a test which is worthless when evaluated scientifically.

The Validation of Personality and Interest Tests The validity of personality and interest tests for employment purposes must be determined by statistical analysis, in the same way as the validity of other tests must be determined. Because of the possible faking of such tests, however, one cannot *really* rely on their validity unless it has been determined by the follow-up method, and unless it has been double-checked by cross-validation on a second "hold-out" group.

Since the responses of applicants to such tests may be different from what their responses would be if they were later being tested as "present employees," it is important that such tests be validated with the *kinds* of responses that *applicants* typically give. It is possible that the validation of tests under such circumstances may result in little or no relationship with the criterion of job success. If this turns out to be the case, then it is better to face this reality than to use the test on the assumption that it is a useful device when it really is not. But it is also possible that the responses given by applicants *may* be related to the criterion. If this is the case, the test can be used for prediction whether the responses are, or are not, in all respects truthful responses.

A study by Tiffin and Phelan[16] illustrates this approach. A metal-parts factory had been giving the Kuder Preference Record (Vocational Form C-H) to applicants for hourly-paid jobs for a period of 14 months.

---

[15] R. Stagner, "The Gullibility of Personnel Managers," *Personnel Psychology,* 11 (1958), 347–352.

[16] J. Tiffin and R. F. Phelan, "Use of the Kuder Preference Record to Predict Turnover in an Industrial Plant," *Personnel Psychology,* 6 (1953), 195–204.

The test results, however, had not been used in the employment process, and it was possible to conduct a follow-up study covering nearly 1,800 men who had been employed during that period. The analysis was made against a criterion of job tenure of the men who had been tested, 1,109 of whom were still employed at the time of the study. Of the remaining 684 who had left, 450 were "voluntary" quits. The tests of the 1,109 present employees and the 450 voluntary quits were randomly assigned to two groups. Two-thirds were used to form a "primary" group for item and subscore analysis; the remaining one-third were set aside as a "hold-out" group for later cross-validation. From the primary group two "extreme" criterion groups were selected; 250 present employees with 11–14 months' tenure formed a "long-tenure" group, and 200 voluntary quits, who had worked for three months or less, formed a "short-tenure" group.

Two analyses were made for these two criterion groups. The first consisted of an item analysis; 76 items having a difference between the two groups of 10 per cent or more were retained to form a new scoring key. The second analysis was made on the basis of the correlations of scores on the several subtest components with the tenure criterion groups. Four of the scales gave evidence of a significant relationship with tenure. The "Outdoor" and "Mechanical" scales were positively correlated with tenure, while the "Persuasive" and "Clerical" scales were negatively correlated. These scales were incorporated in a composite score formula with appropriate, statistically determined weights.

Both the special scoring key and the composite scores (based on subtests) were then tried out with the hold-out group to see to what extent they would differentiate that group on the basis of tenure. For this purpose that group was divided into a "long-tenure" group (those who stayed more than three months) and a "short-tenure" group (those who stayed three months or less). The relationships are shown in Fig. 7.1. This shows, for the two methods of scoring (special scoring key and composite scores), the relationship between scores and tenure, specifically the per cent of short-tenure employees for each test-score group. Both scoring systems were found to discriminate to an appreciable degree. However, the special scoring key differentiated better, but it was not possible in this study to determine whether the difference between the two methods was statistically significant.

With this type of cross-validated evidence, a company could use the test in question with considerable confidence that it would really be useful in screening out before employment many persons who would most likely be short-tenure employees.

In a different context, Miner[17] also developed a special scoring key for the Kuder Preference Record. His concern was in the area of manage-

---

[17] J. B. Miner, "The Kuder Preference Record in Management Appraisal," *Personnel Psychology,* 13 (1960), 187–196.

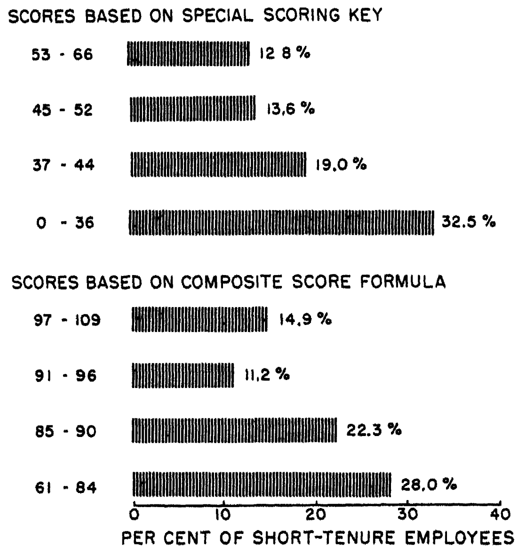


Fig. 7.1. Per cent of short tenure employees by test-score categories based on special scoring key and on composite score formula as used with Kuder Preference Record. (Adapted from Tiffin and Phelan, *op. cit.*)

ment appraisal. His procedure will be described briefly. First he selected 34 items that were tentatively considered as being appropriate for a "supervisory" scale. These were submitted to 20 judges who were asked to select those responses "for which you feel a man who liked to supervise and direct the behavior of others more than anything else would indicate a preference." Seven items were eliminated because the judges did not agree on the relevant "supervisory" response. Tests of 67 management personnel were then "scored" on the remaining 27 items, and the cases were split into a "high" and "low" group at the median of these scores. Response frequencies within these groups were obtained for all these items, and the items that did not *clearly* differentiate between the high and low groups were eliminated; four items were then eliminated, leaving 23 items in the final supervisory scale.

The test was given to 464 management personnel in the company (an integrated petroleum company). One analysis of the test data consisted of a comparison of the scores for research management personnel (n = 65) as contrasted with managers of major company units (n = 50). Of the various scales on the Kuder test, these two groups differed significantly on three, one of them (the "supervisory" scale) being the one described above. The mean scores of the two groups on these three scales are given below.

Interest area on Kuder test	Mean score: Research management personnel	Mean score: Managers of major units
Scientific	52	41
Persuasive	39	46
Supervisory	29	36

It can be noted that research managers tend to be much less interested in persuasive and supervisory matters, and more interested in scientific affairs, than managers of operating units of the company.

The need for rigorous statistical validation of personality and interest tests is reflected in the results of a study with the Rorschach Test. Kurtz [18] reports an attempt to validate this test in the selection of life insurance sales managers. The responses of successful and unsuccessful groups of managers were item-analyzed to derive a special scoring key to differentiate between the two groups. The key thus derived correctly classified 79 of the 80 individuals comprising the two original groups. The necessity of cross-validating a key of this sort [19] on other groups, which were not used in establishing the key, was shown by the fact that when the key was tried with different groups, it did not differentiate better than could be accounted for by chance. A similar study of the Rorschach by Cox [20] gave results on original groups that suggest the usefulness of this test for the selection of salesclerks. The necessary step of cross-validation on new criterion groups, however, was not reported in this study.

Examples of Personality and Interest Tests in Use A few test validation studies with personality and interest tests will be cited as examples. Maher [21] included several tests in an experimental battery of tests with 45 newspaper writers and 44 newspaper advertising space salesmen. The Study of Values test [22] was used with both groups; it is intended to reflect the level of the testee's personal value system in each of six areas (shown in Table 7.4). In addition, the Gordon Personal Profile was used with the newspaper writers. The validity correlations with rating criteria are given in Table 7.4. Two of the tests were statistically significant for the writers, three for the salesmen. The salesmen, for example, had strong values in aesthetics and political affairs, low values on the religious scale.

---

[18] A. K. Kurtz, "A Research Test of the Rorschach Test," *Personnel Psychology*, 1 (1948), 41–51.

[19] E. E. Cureton, "Validity, Reliability, and Baloney," *Educational and Psychological Measurement*, 10 (1950), 94–96.

[20] K. J. Cox, "Can the Rorschach Pick Sales Clerks?" *Personnel Psychology*, 1 (1948), 357–363.

[21] H. Maher, "Validity Information Exchange," 16-01, 16-02, *Personnel Psychology*, 16 (1963), 71–77.

[22] *Study of Values: A Scale for Measuring the Dominant Interests in Personality*, G. W. Allport, Philip E. Vernon, and Gardner Lindsey, Copyright Houghton Mifflin Co.

Another example, reported by Robbins and King,[23] was a follow-up study with a group of male employees of a large manufacturing company who had taken a special sales training program from 9 to 11 years before. At the time of the study, they were categorized in one of six levels in the management sales hierarchy. The Kuder Preference Record and the Bernreuter Personality Inventory were given to the men, and their scores were correlated with the hierarchical ranks of their positions. (This is a "job group" type of criterion, level in the hierarchy presumably reflecting generally their relative ability levels as demonstrated over the years). The following subtests were found to be significantly correlated with their levels:

Kuder: Mechanical, computational, persuasive, artistic, social service (negative correlation), and clerical.
Bernreuter: B1-N (well-balanced emotionally), B3-I (Extroverted), B4-D (Dominant), and F1-C (Self-confident).

Table 7.4

VALIDITY COEFFICIENTS OF CERTAIN TESTS ADMINISTERED TO NEWSPAPER WRITERS AND ADVERTISING SPACE SALESMEN

Test	*Newspaper writers*	*Advertising space salesmen*
Study of Values		
Theoretical	.02	−.01
Economics	.09	.11
Aesthetic	−.21	.42***
Social	−.19	−.12
Political	.19	.45***
Religious	−.33**	−.31**
Gordon Personal Profile		
Ascendancy	.33**	
Responsibility	09	
Emotional Stability	.20	
Sociability	−.23	

* *From Mayer,* op. cit. ** *Significant at .05 level.* *** *Significant at .01 level.*

Izard[24] used the Edwards Personal Preference Schedule in an investigation of the personality characteristics of engineers. In comparing test scores of 81 experienced engineers with corresponding results from 750 male liberal arts students, he found that the engineers were higher

[23] J. E. Robbins and D. C. King, "Validity Information Exchange," 14-02, *Personnel Psychology,* 14 (1961), 217–219.

[24] C. E. Izard, "Personality Characteristics of Engineers as Measured by the Edwards Personal Preference Schedule," *Journal of Applied Psychology,* 44 (1960), 332–335.

on Achievement, Deference, Order, Dominance, and Endurance. They were lower on Affiliation, Intraception, Succorance, Abasement, and Nurturance. Izard concludes from this investigation—which corroborated previous findings of other investigators—that the Personal Preference Schedule is helpful in explaining personality differences such as those found between engineers and non-engineers.

With respect to engineers, Dunnette *et al.*[25] (as well as other investigators) report differences in the patterns of interests between different types of engineers. In this particular study, they administered the Strong Vocational Interest Blank (as well as other tests) to various groups of engineers. With certain groups they used four special scoring keys that previously had been developed for use with engineers, these being: Research, R; Development, D; Production, P; and Sales, S. They found that, with groups of engineers of these four types, these scoring keys generally differentiated among them, as indicated below:

	Mean scores on four scoring keys			
Group of engineers	R	D	P	S
Research	.59	.54	.36	.41
Development	.54	.57	.44	.48
Production	.48	.51	.53	.53
Sales	.46	.47	52	.60

Each group scored highest on the scoring key that was relevant to the group. On the basis of these and other data, these investigators point out that the interest patterns of different types of engineers are somewhat different. Information about the individual interest patterns of individuals would then be useful in mapping their occupational careers.

Validity of Personality Tests for Employment Before ending the discussion of personality tests, it would be desirable to indicate in some way the extent to which they have been shown to have validity for employment purposes. A survey by Ghiselli and Barthol[26] brings together, and summarizes very effectively, a total of 113 studies dealing with the validity of personality tests for employment purposes. About 40 per cent of the material in the survey was from unpublished sources. The studies included in the survey were restricted to those in which the criterion used was some index of job proficiency, such as production records or ratings by supervisors. An attempt was also made to include only those studies in which the scoring key for the personality test was developed independently

---

25 M. D. Dunnette, P. Wernimont, and N. Abrahams, "Further Research on Vocational Interest Differences Among Several Types of Engineers," *Personnel and Guidance Journal* (January, 1964), 484–493.

26 E. E. Ghiselli and R. P. Barthol, "The Validity of Personality Inventories in the Selection of Employees," *Journal of Applied Psychology*, 37 (1953), 18–20.

of the group for which the validity coefficient was reported; no mention was made, however, of the extent to which the cases actually were limited to this type.

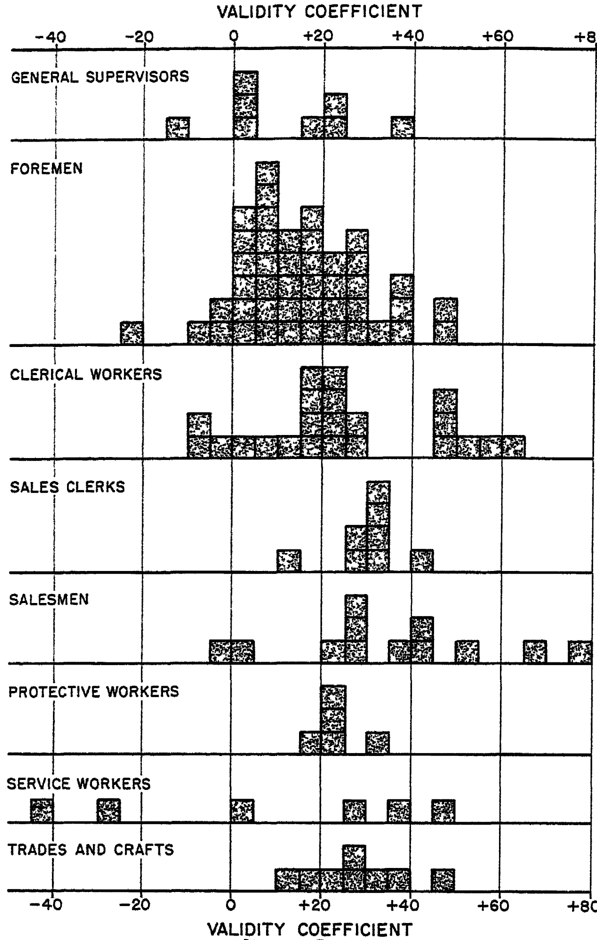


Fig. 7.2. Distribution of validity coefficients of personality inventories for various occupational groups. Decimals are omitted from coefficients. (From Ghiselli and Barthol, *op. cit.*)

The results are shown in two ways. Fig. 7.2 gives the distribution of validity coefficients of the tests for various occupational groups, and Table 7.5 lists the weighted mean validity coefficients for the occupational groups. On the basis of these results, personality tests seem to be potentially more useful for selection of clerks, salesclerks, salesmen, protective

workers, and trades and crafts, than for the other occupational groups. It can be seen, however, that in some groups the data are based on only a few cases, as with the protective workers and service workers. It can also be seen (from Fig. 7.2) that the results are much more consistent for some occupations than for others. For example, the results are more consistent for salesclerks, protective workers, and trades and crafts, than for salesmen and service workers.

Some of these results look very encouraging, but a note of caution must be added. Although not so stated in the report, it is probable that many, if not most, of the studies covered by the survey were studies in which *present employees* had been tested rather than *applicants*. Because of the limitations mentioned before, it is very possible that the validities in some of these cases would be *lowered* if they were established on the basis of *applicant* groups. The results given in Fig. 7.2 and Table 7.5 therefore probably should be considered as over-estimates of the validities that would actually occur in true-life employment situations.

Table 7.5

WEIGHTED MEAN VALIDITY COEFFICIENTS OF PERSONALITY INVENTORIES FOR VARIOUS OCCUPATIONAL GROUPS

Mean r	*Total No of Cases*	*Total No. of r's*	*Occupation Group*
.14	518	8	General Supervisors
.18	6,433	44	Foreman
.25	1,069	22	Clerks
.36	1,120	8	Salesclerks
.36	927	12	Salesmen
.24	536	5	Protective Workers
.16	385	6	Service Workers
29	511	8	Trades and Crafts

Source. *Ghiselli and Barthol,* op. cit.

Other Uses of Personality and Interest Tests This discussion of personality and interest tests deals primarily with their use for employment purposes. It should be kept in mind, however, that such tests have other uses as well. In fact, in many, if not most, other situations they would be less likely to have the limitations that have been mentioned above. For example, in vocational guidance or clinical counseling the counselee most typically is genuinely interested in finding out something about himself, and usually he will be relatively honest in responding to the test items. Even in an industrial setting there may be circumstances where the individual being tested would be motivated to be as honest as possible in taking a personality or interest test. This might be the case, for example, with a person who has been hired, but has not been specifically placed on

a particular job, or with a person who is being counseled by someone in the company about a personal problem or a job placement problem. In such situations personality and interest tests may be very useful devices.

The point is reiterated here, however, that it is particularly important that such tests be thoroughly validated if they are to be used for employment purposes.

8

Achievement Tests

Achievement tests, as we have noted, are used to measure the level of proficiency of individuals in some work activity or subject matter. In business and industry such tests are used primarily for the following purposes: (1) the employment of people for jobs for which experienced workers are sought, (2) the transfer and/or promotion of present employees, and (3) in connection with training (as in identifying areas for which additional training is desirable and measuring the results of training that has been given).

Job achievement tests may be of several forms, such as standardized oral questions, written tests with objective questions, or job samples. Some job achievement tests must be given individually; others may be readily administered to groups of individuals. In comparing achievement tests with aptitude tests (as used for employment purposes), achievement tests generally measure how much actual job proficiency the applicant is able to demonstrate at the time he seeks employment, whereas aptitude tests measure the latent capacity to develop proficiency.[1]

---

[1] Those interested in the development of achievement tests are referred to Dorothy C. Adkins, *Construction and Analysis of Achievement Tests* (Washington, D.C.: U.S. Government Printing Office, 1947).

Validity of Achievement Tests The validity of achievement tests is usually established on the basis of *content* validity (in which the appropriateness of the content is determined by "expert judgment"), or on the basis of *concurrent* validity (in which the test is evaluated by showing how well the test scores correspond with measures of some concurrent criterion). The validation on the basis of concurrent validity is by all odds the preferable method, when circumstances make it possible. When this method is used, it is most typically the practice to use two or more job-groups as the criterion groups, such as a group of journeymen in a trade, and a separate group of apprentices in the trade, for purposes of comparison.

There are various types of achievement tests, but they can generally be grouped into three classes: namely, job sample (or performance) tests, written tests, and oral tests. Some examples of these will be illustrated.

JOB SAMPLE TESTS

The job sample (also called a job performance test, a job miniature, or a work sample) is an achievement test that consists of trying out the individual in a test situation that reproduces all, or an important sampling of, the actual operations that the job itself requires. Such a test usually utilizes equipment such as that used on the job. If the equipment actually used on the job is hazardous, simulated equipment is sometimes constructed which eliminates the hazards of the usual machines. A scoring procedure is developed, and norms of experienced and inexperienced workers are usually obtained in the test situation as a basis of evaluating scores of persons taking the test, such as applicants.

A Miniature Punch Press An example of this approach is the miniature punch press illustrated in Fig. 8.1, which is a replica in all essential features of a small industrial punch press. It differs from a real press in that the punch is located in a vertical bearing and is held down only by a spring. This feature prevents the punch itself from descending when an obstacle is encountered. When this occurs, the punch remains stationary while a mechanical counter records an error or mispunch. The test is administered by having an applicant put through the press 200 prepunched pieces of galvanized sheet iron. The time required to feed these pieces is recorded by means of a stop watch. During the test period the mechanical counter records the number of errors or mispunches. The test thus results in simultaneous time and error scores.

The validity of this test has been studied by comparing the average performance of different groups of persons to whom the test had been given. The curves in Fig. 8.2 summarize the results obtained for three

Fig. 8.1. Applicant being given the Miniature Punch Press Test.

specific groups. These curves show the relation between mispunches or errors in punching 200 plates and the time in minutes required to punch the 200 plates. As might be expected, the errors decrease as the time increases for all three groups of persons upon whom these results are based. It is interesting to note, however, that the curve for the students and the one for the insulation stripping machine operators are almost identical mathematically and simply represent different segments of the same curve.

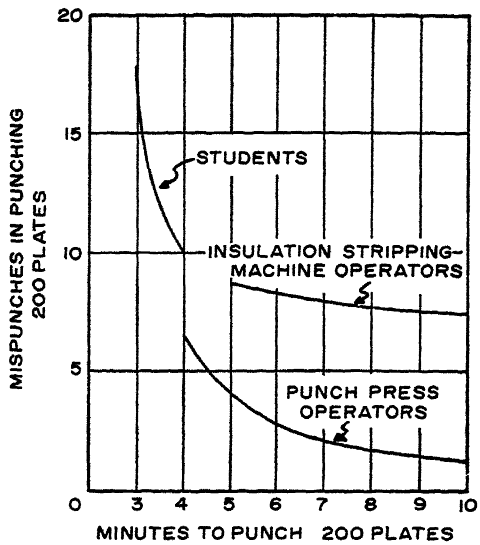


Fig. 8.2. Relation between speed and accuracy for three groups of subjects in operating a small punch press.

This suggests that the students, though punching much more rapidly and hence making many more errors, are not significantly different from the insulation stripping machine operators in genuine ability. The employees in this latter group, though offering no experience in the operation of a punch press, were industrial employees and, as such, were more accustomed to the need for careful, slow operation of any machine. The curve for the industrial punch press operators, however, is markedly different from that of either the students or the insulation stripping machine operators. For any given speed of operation the punch press operators were more accurate than either of the other groups, and for any given level of accuracy, punch press operators were more rapid.

From a knowledge of the speed and accuracy of a given applicant and a comparison of this information with the data graphed in Fig. 8.2, it is possible to determine the status of the applicant in comparison with the corresponding test performance of persons who are known to be experts on this job.

The method illustrated by the miniature punch press has been utilized in selecting employees for numerous other jobs. For example, in hiring persons for such jobs as packaging, inspecting, and certain types of machine operation, it is often extremely helpful for the personnel manager to obtain a sample of the quality and speed of work that the applicant is able to perform.

A Performance Test for Fork-Lift Operators An example of a very different type of job sample test is one that has been developed by the Philadelphia Quartermaster Depot for fork-lift operators.[2] A fork-lift truck is used for moving materials in a warehouse or around an industrial plant. The test consists of driving a loaded fork-lift truck around a standard driving course as shown in Fig. 8.3. The "walls" and the

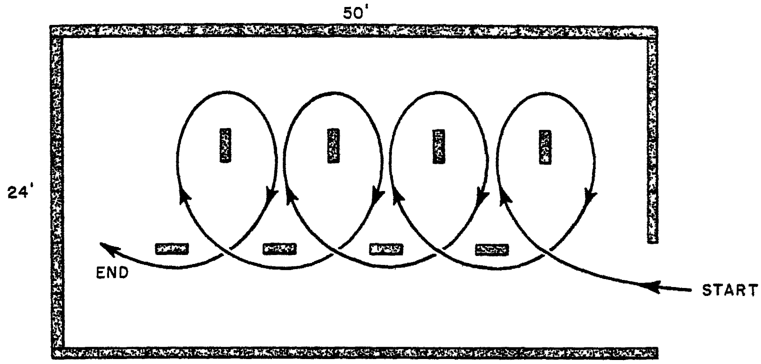


Fig. 8.3. Standard driving course diagram for performance test for fork-lift operator. (Test prepared by U.S. War Dept., Office Quartermaster General, Philadelphia Quartermaster Depot, Testing Station.

obstacles are constructed from pallets 32 by 40 inches, painted yellow to contrast with the floor. Without going into detail, it should be pointed out that the equipment used and instructions should be standardized. The person giving the test has a check sheet on which he records his observations. This check list includes a listing of 42 poor operating techniques such as "Did not start in low gear," "Started jerkily," "Scraped side walls of tires," "Lowered load too quickly," and "Number of pallets displaced." Scoring is based on the number of errors made, as recorded on the check list.

Although it might appear that many jobs would not lend themselves to job sample tests, it can be seen from an example such as this that the use of imagination and inventiveness can result in the development of tests that can be objectively scored for a wide variety of jobs.

The Job Sample in Clerical Work Another area in which the job sample method of testing has resulted in very satisfactory results is the selection of employees for stenographic, clerical, and secretarial positions. One of the most widely used tests of this type is the series known as the

[2] This test was prepared by the U. S. War Department, Office Quartermaster General, Philadelphia Quartermaster Depot, Testing Section.

Blackstone Stenographic Proficiency Tests.[3] This series includes a stenography test and a typewriting test. The stenography test is made up of seven parts, namely, knowledge of English grammar (which includes punctuation, capitalization, and spelling); syllabification; office practice; alphabetizing; abbreviations; knowledge of business organization; and ability to take dictation and shorthand and to make the correct transcription. Separate norms are available for each of these seven parts so that in hiring an applicant for a given job it is possible to match the test results with a job analysis of the work to be performed. The second part of the Blackstone series consists of a typewriting test. This part requires the applicant to copy a standard page of typewritten material. The result may be scored separately for speed and accuracy and the norms available permit a rapid determination of the exact level of proficiency possessed by an applicant. The Purdue Clerical Adaptability Test [4] can also be used effectively as an achievement test.

A standard test used in measuring proficiency in the use of a typewriter is the Thurstone Examination in Typing.[5] A part of this test consists of a page of corrected copy that the applicant is asked to type, making the indicated corrections. A sample of the material to be copied is shown in Fig. 8.4. The ability to copy corrected material of this kind, though

The typical business man is an optimist. For him, the future
is full of possibilities that have never been realized in the past.
He is not, however, a day-dreamer, but one ~~whose~~ who uses his imagination ~~is used~~ in
~~setting up~~ formulating purposes which lead to immediate action. His power~~s~~ of
execution ~~and planning~~ often surpasses that of his imagination, and
he is ~~often~~ frequently surprised to see ~~have~~ realized his vision in less time
than he had even dared to hope.

Fig. 8.4. Part of Test 1 of the Thurstone Typing Test. The person tested is required to copy this material, making the indicated corrections.

frequently called for in many stenographic positions, seems to be quite different from the ability to make an exact copy of material requiring no corrections. For example, in one employment office a girl who had recently won a state contest for speed in typing was unable, when tested with the Thurstone test, to obtain even an average score in typing this corrected copy. Apparently her ability consisted of a strictly mechanical

[3] *Blackstone Stenographic Proficiency Tests* (New York: Harcourt, Brace & World, Inc., 1932).

[4] See Appendix D.

[5] *Thurstone Employment Tests, Examination in Typing* (New York: Harcourt, Brace & World, Inc., 1922).

and automatic method of "copying the copy." Such ability might be very useful in winning state contests, but it is of little use in a business situation because one is seldom asked to copy material that is perfect at the outset. A test of the Thurstone type, therefore, is a much better measure of the kind of ability called for in a business or industrial job than is a test that requires only straight copying of material.

An excellent work sample test of transcription and typing has been recorded on standard phonograph records by Seashore and Bennett.[6] Letters are dictated from records at varying speeds. The applicant takes the material in shorthand and later types the letters in standard form. The score on the test is determined by the appearance and accuracy of the final typed letters.

WRITTEN ACHIEVEMENT TESTS

Most written achievement tests are "tailor made" for the particular situation, such as in connection with training programs that are directed toward specific job-training purposes. In such contexts, they should be developed with acceptable test-development practices in mind. In addition to such tailor-made tests, however, there are a few standardized achievement tests that have been developed for more general use.

For example, the Purdue Vocational Tests include tests for certain trade areas such as machine shop practice, electricity, welding, sheet metal work, and carpentry. (These tests are listed in Appendix D.) To illustrate the nature of such tests, however, four items will be given from one of the tests in this series, the Purdue Test for Machinists and Machine Operators.

> In each of the multiple-choice statements listed below there are four possible answers, but only one is correct. Read each statement carefully before making your choice of answers.
>
> By tolerance in machine work is meant (*a*) the allowance, as for the oil film of a bearing, (*b*) the amount of variation either above or below a certain basic measurement that will be acceptable, (*c*) the amount of stock left for polishing, (*d*) the amount of stock left for grinding ( )
>
> If it fits, the wrench which will probably do the least harm to the corners of a nut is (*a*) an adjustable, (*b*) an alligator, (*c*) an open end, (*d*) a socket wrench ( )
>
> One of the best bearing metals contains antimony, tin, and copper. This metal is called (*a*) bronze, (*b*) brass, (*c*) babbitt, (*d*) lead ( )
>
> Standard taper sizes are designated by (*a*) decimals, (*b*) fractions, (*c*) numbers, (*d*) letters ( )

[6] H. G. Seashore and G. K. Bennett, *The Seashore-Bennett Stenographic Proficiency Test* (Psychological Corporation, 1946).

The test covers such topics as hand tools, bench tools, bench work, lathe, milling machine, shaper, planer, and drill press. The test is arranged so that subscores may be readily obtained for bench operations and for the operation of the grinder, planer, lathe, and milling machine. On the basis of odd-*vs.*-even item correlations, the reliability of the various scores was as follows.

Bench operations	.87
Grinder	.90
Planer	.89
Lathe	.80
Milling Machine	.85
Total Score	.96

ORAL ACHIEVEMENT TESTS

Oral achievement tests are used in some situations, usually in connection with various trades. When used for trades these tests are commonly called *oral trade questions.* Such questions are convenient to administer and simple to interpret. The most extensive research with this type of test has been carried on by Stead, Shartle, and associates. The general procedure followed by this group has been described in detail in their book, *Occupational Counseling Techniques.*[7] Since the many trade questions developed are in use by the various state employment offices, and since the publication of the actual questions used would cause them to lose much of their value, Stead and Shartle publish only a few sample questions to illustrate the method by which the validity of the questions has been determined. This procedure is sufficiently important to the general subject of test validation to warrant a somewhat detailed discussion.

Validation of Oral Trade Questions The general procedure used by the U.S. Employment Service in validating these questions consisted of finding for each trade a few questions, usually 15, each of which was answered correctly by a large percentage of successful journeymen in that trade, but was answered correctly by a definitely smaller percentage of apprentices in the trade, and by a still smaller percentage of employees in related occupations. When a set of questions for a given trade has been obtained in this manner, it is a simple matter to determine whether an applicant who claims to possess journeyman abilities for that trade actually is able to answer these questions as well as known tradesmen have answered them. The differentiation obtained by such a set of oral trade questions for asbestos workers is shown in Fig. 8.5. The expert asbestos workers vary over a range of 7 to 15 correct answers;

[7] W. H. Stead, C. L. Shartle, *et al., Occupational Counseling Techniques* (New York: American Book Company, 1940).

FORM I

Score	Expert Asbestos Workers (50 subjects)	Apprentices and Helpers (25 subjects)	Related Workers (25 subjects)
15	xxxxxxxx		
14	xxxxxxxxxxxxxxxxx*		
13	xxxxxxxxxxxx		
12	xxxxxx	xx	
11	xx		
10	x	x	
9	x		
8	xx	xx	
7	x	x	
6		xxxxxxxx*	
5		xxx	xxxx
4		xxxx	xx
3		xx	xxx
2		x	xx
1		x	xxxx*
0			xxxxxxxxxx

*Median score

Fig. 8.5. Distributions of scores for expert asbestos workers, apprentices and helpers, and related workers on a 15-question oral trade test.

apprentices and helpers, of zero to 12; and related workers, of zero to 5. Thus, an applicant who is able to answer correctly only 6 or fewer of the 15 questions is very unlikely to be an expert asbestos worker, regardless of what he may say or feel about his own skill. Even if he answers up to 12 items correctly, he is probably not highly proficient in his trade, because the majority of men known to be experts were able to answer 13 or more of the questions correctly.

Factors Affecting the Validity of Trade Questions The success with which such a set of trade questions will differentiate among applicants possessing different degrees of skill depends upon the very careful selection of questions comprising the test. This selection can be made only by starting with many more questions than will finally be retained and by eliminating every question that does not actually show a differentiation among the several groups tested. Likewise, if the test is to be used on a country-wide basis, the preliminary validation work must be carried on in the several geographical areas where the test is later expected to function. The following question for a roofer illustrates the necessity for considering geographical factors in the validation of the question.

Q. What type of asphalt is used on a flat roof?
A. Flat (F) (low melt).[8]

---

[8] *Ibid.*, p. 38.

Experts west of the Mississippi River answered this question with unusual consistency, but there tended to be a reduction of consistent answers by experts when the validation took place eastward. Therefore this question could not be used in a test that was intended for countrywide use. In general, questions that could be answered simply, preferably by one or two words, were found to be better than those that called for more extensive answers. Often the best questions were those that could be correctly answered only by terms peculiar to the trade in question, even though such terms might not be found in a dictionary.

Trade questions of this general type have been prepared for most of the standard trades and are in use in many state employment offices.

Although most oral trade tests include questions to which short, concise answers are to be given, true-false questions can also be used. Examples of true-false questions covering the job of heater in a steel mill are given in Table 8.1.

Table 8.1

TRADE QUESTIONS IN TRUE-FALSE FORM COVERING JOB OF HEATER IN A STEEL MILL

1. The furnace checkers are always located directly below the car bottom	T	F
2. The gas flow and air flow must always come from opposite sides of the furnace	T	F
3. The only way to tell if the stock gas is burning is to look at the top of stock	T	F
4. The purpose of the stock gas is to help the furnace draw	T	F
5. In lighting a cold furnace, the rear parts are always the first ones to be opened	T	F

Industrial Uses of Oral Achievement Tests Although oral achievement tests apparently have not been adopted very widely by industry, this testing technique seems to offer considerable promise for measuring achievement. In addition to differentiating different *job* groups (such as journeymen and apprentices), such tests also can differentiate rather well different levels of ability *within* a job. This was indicated, for example, in a study by McCormick and Winstanley[9] of machinists in a shipyard.

Ninety-eight machinists took a 21-item oral test that included questions such as those shown in Table 8.2. The 21 items in this test had been identified from an original test of 95 items as being the ones that were most discriminative. The machinists were divided into three approximately equal groups (A, B, and C) on the basis of ratings by their super-

[9] E. J. McCormick and N. B. Winstanley, "A Fifteen-Minute Oral Trade Test," *Personnel*, September, 1950.

Table 8.2

TYPES OF QUESTIONS INCLUDED IN ORAL TRADE TEST FOR SHIPYARD MACHINISTS

1. *Q.* What material is used in the manufacture of lapping plates?
   A. Cast iron.
2. *Q.* How many turns of a handle are required to make a half turn of the most common index head?
   A. 20.
3. *Q.* What is the meaning of the Brinnel test?
   A. A method of determining the hardness of metal.
4. *Q.* For what purpose would you use transfer calipers?
   A. Used to determine size of recesses and places where the legs of the calipers must be moved to get them out.

Source: *McCormick and Winstanley*, op. cit.

visors on "over-all" job performance. Fig. 8.6 summarizes the results, by indicating the per cent in each criterion group (A, B, and C) whose scores were 13 and above; it shows that the test differentiates quite clearly among the three criterion groups. It was suggested that such a test might well be used for determining the need of individuals for training, for upgrading, for placement or transfer, for selection, for measuring level of ability, and for use in handling grievances that deal with questions of promotions.

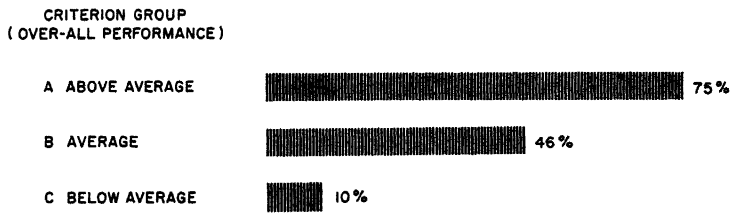


Fig. 8.6. Per cent of machinists in each of three criterion groups scoring at or above critical score of 13 on oral trade test. (Adapted from McCormick and Winstanley, *op. cit.*)

USES OF ACHIEVEMENT TESTS

Some of the uses of achievement tests have been mentioned or illustrated briefly above. A few additional illustrations will be given here.

Employment One of the most important uses for achievement tests is in the employment of applicants. When a company is hiring tradesmen it is often difficult to determine from an interview whether an applicant has adequate knowledge of the trade in question. In this situation, a

paper-and-pencil trade information test very quickly and efficiently shows how much the applicant knows about the trade. An example of this use of a trade information test is shown in Fig. 8.7. The Purdue Trade Information Test in Engine Lathe Operation was administered to 60 men. Thirty of these were journeyman machinists, and the remaining 30 were high school graduates who had had at least one year of vocational machine shop instruction. As Fig. 8.7 shows, the test very sharply differentiated between the journeymen and the boys right out of school.

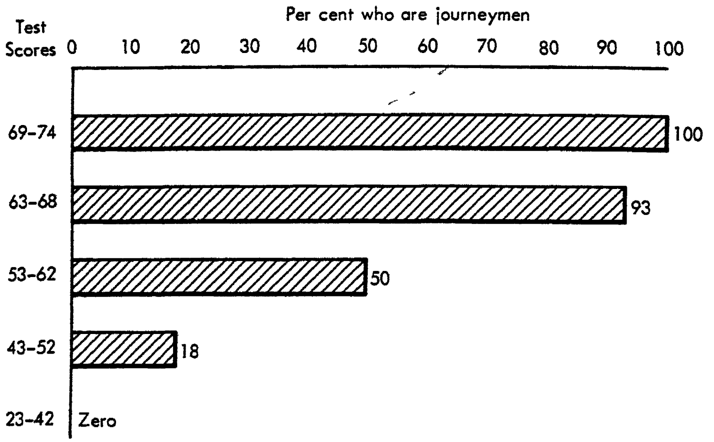


Fig. 8.7. Per cent of men who were journeymen machinists as a function of score made on the Purdue Trade Information Test in Engine Lathe Operation.

Similar results were obtained with the Purdue Trade Information Test in Welding. These results are shown in Fig. 8.8. Here again the trade information test clearly helped to separate the journeymen from boys right out of the vocational school.

Another rather typical employment problem where trade information tests are useful is in the selection of apprentices for the various skilled trades. This is an important problem because of the time and expense involved in apprentice training. Since many applicants for apprenticeships have had high school courses in areas related to the trade involved, achievement tests have been found to pay substantial dividends in the selection of men for the apprenticeships. In such a case an achievement test covering a certain subject area is useful in the prediction of continued development in the subject area in question. (In one sense, such tests are then used more as aptitude tests than as achievement tests.)

The desirability of using a test in such a circumstance is illustrated

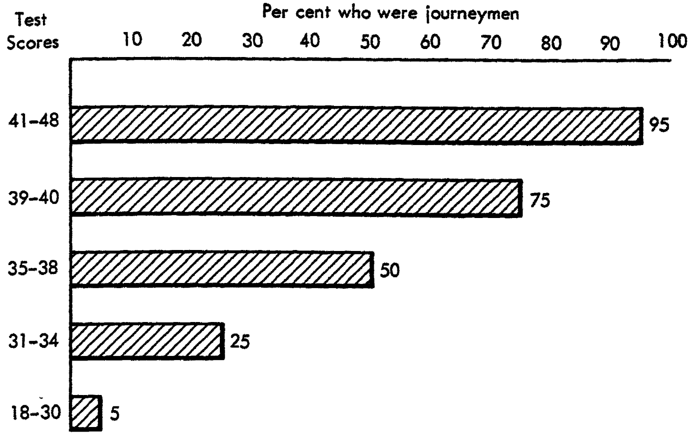


Fig. 8.8. Per cent of men who were journeymen as a function of score made on the Purdue Trade Information Test in Welding.

in Fig. 2.4 in Chapter 2 on individual differences. This shows that among 112 applicants for machine shop apprentice training a considerable number made a score of less than 40 (number of items correct) and a scattered few made a score of less than 20 on the machine shop achievement test. Although most of these applicants had had at least one semester of machine shop instruction, and many had had much more than this amount of training, 25 per cent of them did so poorly on the machine shop achievement test that they were below the lowest 10 per cent of students in vocational classes. It was clear, therefore, that among these applicants were a considerable number who would be very unlikely, even with a prolonged training, to become expert machinists. The company concerned set 90 items correct as the critical score for hiring for this job and was able, by this process, to select for apprenticeships boys who made very rapid strides in developing the necessary skill to become expert machinists.

Transfer and Promotion of Employees The problem of transfer and promotion is one that continually confronts every operating management. The use of appropriate achievement tests can aid in identifying individuals who have the skills or knowledges required for other jobs. Note that we mentioned *appropriate* achievement tests. The fact of a person being proficient on his *present* job often gives no assurance of abilities for *other* jobs. In other words, an achievement test used for transfer or placement purposes must be valid *for the job for which the individual is being considered.*

The use of tests for transfer and promotion purposes should sup-

plement, rather than replace, other considerations such as seniority, being "in line" for the job, and the considered judgments of management representatives regarding the qualifications of the individual in question.

Training Many companies are devoting more and more time to the systematic training of both new and old employees. The need for such training is a natural result of the continuous technological changes oc-

Table 8.3

PART OF AN ACHIEVEMENT TEST COVERING INFORMATION THAT SHOULD BE KNOWN TO EMPLOYEES SEEKING TRANSFER OR PROMOTION

Instructions: On the right is a list of colors. On the left is a list of the materials carried in pipes in this mill. You are to show how well you know the color code by matching each color with the figure or figures which you find before the appropriate materials. Mark your choice in the parentheses at the extreme right. The first one is correctly marked to show you how it should be done. There will be some colors unused. Some others will be used twice.

Materials	*Colors*	
1. Stabilized Gas	Aluminum	( ) (6)
2. Steam	Black	( ) ( )
3. Hot Water	Brown	( ) ( )
4. Cold Water	Bright Red	( ) ( )
5. Coke Oven Gas	Ceiling Blue	( ) ( )
6. De-oxidized Gas	Dark Purple	( ) ( )
7. Compressed Air	Dark Red	( ) ( )
8. Natural Gas	Gray	( ) ( )
9. Fuel Oil	Green	( ) ( )
10. Sulphuric Acid	Blue	( ) ( )
11. Farval Grease	Pink	( ) ( )
12. Bowser Oil	Olive Green	( ) ( )
13. Gasoline	Orange	( ) ( )
14. Kemp Lines	Lavender	( ) ( )
15. Water Fire Lines	White	( ) ( )
16. Foamite	Yellow	( ) ( )
17. Hydraulic Lines		

curring in modern industry. No matter how well qualified an employee may be today, technological change in methods or processes may require that he be completely retrained tomorrow. The systematic use of technical information tests among present employees furnishes a convenient means of determining those areas in which training is needed. An example of this use of information tests is shown in Table 8.3. This table consists of a set of matching items to determine whether the employees are familiar with the color code used in the plant in which they work. Several other important areas were covered in the tests. The content of the training program that followed the administration of this test was based largely upon the results obtained.

An interesting use of achievement tests in training is described by Taylor,[10] the report dealing with an automotive mechanics course at an Army Ordnance School. The Army personnel going through this training course typically went through three four-week phases of training, adding to a total of 12 weeks. The three phases increased in complexity.

A series of 15 tests was developed for use in the training program, including both written and job sample tests. An analysis of experimental results with these resulted in the selection of five of them for the purposes to be described. These five included an Auto Mechanics Multiple Choice Test, a Tool Usage Film Strip Test, an Engine Ignition Checks Test, a Hand Tools Checks Test, and an Auto Mechanic Experience Check List. From a total group of 807 incoming trainees, the 178 with the highest scores on these tests were identified. These students "skipped" the first phase and entered training at the beginning of the second phase. At the end of the training (in this case 8 weeks), their performance was compared with that of the remaining 629 students who had gone through the full 12 weeks. Performance at the end of training was compared on the basis of final grades in the training course. This comparison, shown in Fig. 8.9, indicates very clearly that those who skipped the first four weeks were, as a group, superior to the others at the end of training, despite the fact that they had had only 8 weeks of training. For this particular group of 178 men, 4,984 man days (13.7 man years) were saved as a result of eliminating the first four weeks of training for them. This reduction in training was made possible, of course, by using the achievement tests for identifying the individuals whose initial level of achievement was highest.

While written achievement tests are used very commonly for measuring present knowledge or achievement in training of people in trade and technical areas, they are also appropriate in other areas, such as in management and supervisory training, in which the emphasis might be on such topics as management policies and procedures, human relations, economics, etc.

Measurement of Vocational Achievement Although our primary interest in achievement tests is concerned with their use in business and industry, it should be pointed out that such tests also have great utility in vocational training work. Every vocational teacher expects to place his students in industrial jobs. The success of these students in such jobs, however, depends largely upon the adequacy of training they have received while students. Teachers in the field of general education have long made use of standardized achievement tests to determine student achievement in the several school subjects and to compare the achievement of students in various school systems, under different types of in-

---

[10] C. W. Taylor, "Pre-testing Saves Training Costs," *Personnel Psychology*, 5 (1952), 213–239.

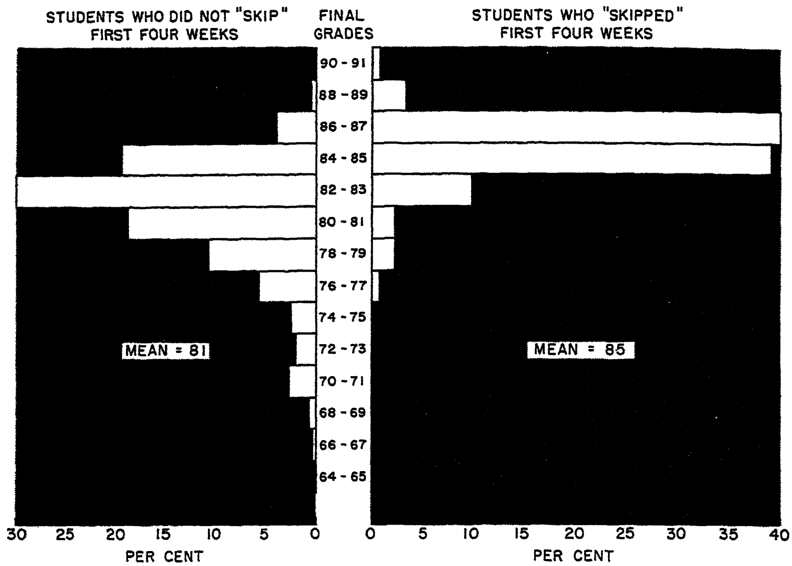


Fig. 8.9. Distributions of final average scores for automotive mechanic students who "skipped" the first four weeks of training and for those who did not. Those who "skipped" were individuals who scored highest on five achievement tests taken at the beginning of training. (From Taylor, *op. cit.*)

struction, and in different geographical localities. Administrators and teachers of vocational subjects are now beginning to make similar use of objective achievement tests, and there is every reason to believe that the judicious use of such tests in this area will be fully as valuable as in the field of general education. The several trade achievement tests listed in Appendix D are ideally suited to this purpose. Since vocational demands differ from one industry to another and from one industrial center to another, it is often wise to build tests that are "tailor-made" for the particular situation. Such tests, when constructed through the joint efforts of vocational teachers, school administrators, vocational co-ordinators, and representatives of the industries concerned, furnish an ideal means of facilitating the co-ordination between industries and school that every community desires to encourage.

9

Performance Appraisal

Performance appraisal is a systematic evaluation of personnel by their supervisors or others who are familiar with their work performance. Such appraisals usually involve the use of forms and procedures that have been developed for the purpose. The recorded evaluations usually become a part of the personnel records of individuals, and are available for use in connection with possible personnel actions such as promotions, transfers, and separations. Performance appraisals are sometimes referred to as employee appraisals, merit ratings, personnel ratings, personnel performance evaluations, as well as by other terms.

THE GROWTH OF PERFORMANCE APPRAISAL

A survey of 400 companies carried out by the National Industrial Conference Board [1] indicated that about one half had employee performance appraisal plans. Although this is not a majority of the companies surveyed, it represents a significant proportion. Despite certain

[1] "Personnel Practices in Factory and Office," *Studies in Personnel Policy,* No. 145, National Industrial Conference Board, 1954.

criticisms that have been made regarding performance appraisal systems, it seems probable that such systems will remain an important part of personnel administration programs of many companies. Such an assumption justifies a rather careful discussion of this topic.

One point should be made clear immediately: although formal performance appraisal systems are of relatively recent origin, the rating of men by supervisors is by no means a new development. Supervisors have *always* evaluated their subordinates, and it is no doubt true that evaluations made in a random, slipshod, and unsystematic fashion, unrecorded and undefended, whether valid or not, have in the past been just as important in influencing personnel actions as evaluations made by current formal evaluation systems. The changes that have come with systematic performance appraisal systems do not involve making evaluations where none existed, but rather involve a shift from informal, unsystematic evaluations to those that are made more systematically, and in a manner that probably results in more comparability from one individual to another. The question, then, is not whether supervisors should rate their employees—this always has been and probably always will be done—but whether the use of a formal merit rating system will increase the value of such ratings both to management and to employees.

PURPOSES OF PERFORMANCE APPRAISAL

Various reasons have been given by company executives for using appraisal systems. In a survey reported in *The Management Record*,[2] 15 different uses were listed. In general, however, these purposes seem to fall into two major categories: "administrative" and "self-improvement" (or clinical). In the second category, the emphasis is upon helping employees to understand their strengths and shortcomings, so that they can have some basis for self-improvement.

A summary of the uses of performance appraisal suggests that the following are the most important.

ADMINISTRATIVE USES OF PERFORMANCE APPRAISALS

Promotion. It is probable that this is the most important administrative use of performance appraisals. It is to the common interests of both management and employees to promote employees into positions where they can most effectively utilize their abilities. A properly developed and administered performance appraisal system can aid in determining whether individuals should be considered for promotions. There

[2] Summarized in "Appraisal of Job Performance," *Studies in Personnel Policy*, No. 211, National Industrial Conference Board, 1951.

is, however, an important distinction to be made in using appraisals for this purpose. Such ratings should differentiate between the *performance* of the individual *on his present job,* and his *potentiality* for performance on a *higher-level job.* The ability to perform effectively on one job does not necessarily give assurance of an employee's potentialities for greater responsibilities. These potentialities need to be evaluated separately.

Other personnel actions. There are occasions in most companies when it is necessary to consider various types of personnel actions such as *transfers, layoffs,* and *discharges.* In some cases such actions may be necessary because of unsatisfactory performance on a present job. In other cases the actions may be necessary because of economic conditions over which the company has no control, or because of changes in the production processes.

Wage and salary administration. In some companies, performance appraisals are used as a basis for granting increases in wages or salaries. In some cases both merit and seniority are used in combination in granting wage or salary increases.

Training. An appropriate performance appraisal system can be useful for training purposes in two ways. In the first place, it can aid in identifying areas of skills or knowledges in which numerous employees are not up to par, thus pointing up general training deficiencies which presumably should be corrected by additional training. In the second place, it can help to identify individuals who may need additional, special training.

Criteria for personnel research. As indicated in Chapter 2, employee evaluations frequently are used as criteria for personnel research purposes such as in test validation.

Employee Self-Improvement For a personnel appraisal system to serve this purpose, it is necessary that each employee know where he stands with his supervisor. This involves an interview between a supervisor or some other company representative and the employee in question. While a discussion of appraisal interviews will be deferred until later, it should be noted here that employee self-improvement must be predicated on some knowledge by the employee himself about his strengths and weaknesses. Such information can be provided to him during an appraisal interview.

CONSIDERATIONS OF MERIT *vs.* SENIORITY

In connection with certain personnel actions such as promotions, wage or salary increases, and layoffs, there is sometimes a question as to the relative importance in such decisions that should be given to merit (performance) as opposed to seniority. This issue is a particularly controversial one in the case of some management-union relationships. In

general, unions have tended to take a dim view of the use of "merit" as the major basis for such action, and, rather, have placed primary emphasis upon seniority. The actual provisions of union contracts regarding the use of seniority and merit for personnel actions vary. Although some such contracts provide that most personnel actions will be based exclusively on seniority, most of them include provisions such as the following: "Promotions within the bargaining unit will be made on the basis of merit, other things being equal," [3] or that seniority shall govern promotions when the "ability, skills, and job performance of applicants are equal." Usually the burden of proof for characterizing relative merit of employees in such cases lies with management.

The extent to which seniority is a factor in various types of personnel actions in non-union companies is summarized in Table 9.1, which is based on a survey among 130 companies.[4] It can be seen that even non-

Table 9.1

THE SENIORITY FACTOR IN DIFFERENT SITUATIONS

Situation where seniority is a factor	*Companies No.*	*Per cent*
Layoffs	116	89
Promotion	107	82
Rehiring after layoff	96	74
Transfer	69	53
Others	7	5

Source: *Adapted from "Seniority Systems in Non-union Companies,"* Studies in Personnel Policy, *No. 110, National Industrial Conference Board, 1950*

union companies place a great deal of emphasis on seniority in taking various personnel actions. However, of a total of 110 of these companies (with less than 400 employees each) only four used seniority as a *sole* basis for granting promotions. More typically, various personnel actions were based in part on other considerations. Eighty-two per cent of the companies, for example, considered merit or ability in conjunction with seniority. Other factors that were taken into account by some companies included marital or family status, physical fitness, hardship, attendance, and cooperativeness.

[3] "Appraisal of Job Performance," *Studies in Personnel Policy,* No. 121, National Industrial Conference Board, 1951.

[4] "Seniority Systems in Non-union Companies," *Studies in Personnel Policy,* No. 110, National Industrial Conference Board, 1950.

PERFORMANCE APPRAISAL SYSTEMS

The individual performance appraisal plans of specific companies vary a great deal, but they usually fall into one or another of certain basic types. Some plans embody certain aspects of two or more basic methods. The most important basic systems are listed below:

1. Rating scales.
2. Employee comparison systems.
   a) Rank-order system.
   b) Paired comparison system.
   c) Forced distribution system.
3. Check lists.
   a) Weighted check list.
   b) Forced-choice check list.
4. Critical incident technique.
5. Other methods.

The 5 categories above are all based on essentially different rating procedures. Since each of the various methods has certain inherent advantages and disadvantages, there is no single "best" method; rather, one method may best serve one purpose, and another method may best serve a different purpose.

Rating Scales Rating scales, also referred to as "the chart system," are the most widely used type of performance appraisal system. The basic principle of this method provides for the rating of employees on each of a number of different traits or worker characteristics. There are two primary variations in the manner in which the ratings may be made. In the *graphic* rating scale, a line represents the range of the trait or characteristic, and the rater places a check mark at the position along the scale that he considers to represent the degree of the trait for the employee being rated. An example is shown in Fig. 9.1. The *multiple-step* rating

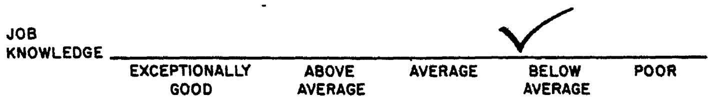


Fig. 9.1. Illustration of graphic rating scale. The rater places a check mark at the position on the scale that represents his judgment of the degree of the trait possessed by the employee being rated.

scale provides for rating on each trait in any one of a number of "degree" categories, as shown in Fig. 9.2.

PROGRESS RECORD

Name________________ Dept.______________ Div ______________ Date__________

Employee's Position__________________________________ Job Class __________

Note This rating will represent in a systematic way your appraisal of the employee in terms of his ACTUAL PERFORMANCE ON HIS PRESENT JOB In the interests of furthering careful analysis, the following suggestions are offered regarding the use of this form

1 Consider only **one** factor at a time
2. Study each factor and the specifications for each grade
3 Review upon completion to see that the rating of each factor applies exclusively to the individual's ACTUAL PERFORMANCE ON HIS PRESENT JOB
4 Comment fully at bottom of page and on reverse side upon any matter which in your opinion needs explanation

PERFORMANCE FACTORS	PERFORMANCE GRADE				
	Far Exceeds requirements of this job	**Exceeds** requirements of this job	**Meets** requirements of this job	**Partially Meets** requirements of this job	**Does Not Meet** requirements of this job
QUALITY OF WORK Accuracy Economy of Materials Economy of Time (his own and others) Neatness Thoroughness	Consistently superior	Sometimes superior	Consistently satisfactory	Usually acceptable	Consistently unsatisfactory
QUANTITY OF WORK Productive Output	Consistently exceeds requirements	Frequently exceeds requirements	Meets requirements	Frequently below requirements	Consistently below requirements
DEPENDABILITY Follows Instructions Judgment Punctuality and Attendance Safety Habits	Consistently dependable	Dependable in most respects	Ordinarily dependable	Frequently undependable	Consistently undependable
COMPATIBILITY Attitude Towards the Company Attitude Towards Supervision Co-operation with Fellow-Employees	Inspires others to work with and assist co-workers	Quick to volunteer to work with and assist others	Generally works well with and assists others	Seldom works well with or assists others	Does not work well with or assist others

COMMENTS ____

Fig. 9.2. Example of a multiple-step rating scale. This is a typical rating scale used in industry.

The rating systems used in various companies differ widely in the number of traits to be rated and in the particular traits used. The variety of traits may be seen from an analysis of the merit-rating charts of eighteen companies [5] as shown in Table 9.2. The companies are referred to anonymously by number across the top row of the table. The number of traits included varies from 21 in the case of Company 1 (first column), down to 4 in the case of companies represented by the last two columns. The median number is 10. In the development of performance appraisal systems, however, the principle should be followed of developing a system that is relevant for the intended purposes, rather than trying to conform to what others are doing. There is, however, some reason to believe that there is a tendency to include provision for fewer rather than more traits in performance appraisal systems.

There are two possible justifications for this tendency. In the first place, there is reason to believe that there are not very many different variables on which it would be relevant to obtain ratings. It is doubtful, for example, that there are, say, 21 really different traits or job behaviors on which it would be useful to obtain appraisals. In the second place, the "halo" effect tends to minimize whatever real differences there are among rating variables. The halo effect will be discussed later in this chapter.

Rating factors. Some evidence that is relevant to this matter comes from a factor analysis of approximately 1,100 ratings selected at random from a plant in which about 9,000 employees had been rated.[6] The rating form used involved ratings on 12 separate traits. These traits and the intercorrelations between the ratings for each pair of traits are shown in Table 9.3. Such a correlational matrix containing correlations in general as high as those in Table 9.3 is sufficient in itself to indicate that a common factor is operating to influence most, if not all, of the ratings. The results of a factor analysis of these ratings quite definitely justify this judgment. As discussed earlier, factor analysis is a statistical technique that reduces a set of measurements (such as test results or ratings) to the minimum number of basic variables or factors that will account for the variations in the original data. The factor analysis of the merit ratings revealed three basic factors, factor loadings for which are given in Table 9.4. In interpreting the importance of the different merit-rating items as they enter into each of the three factors, we should bear in mind that the amount of the factor loading indicates the extent to which that item is related to the factor in question. An inspection of the merit-rating items

---

[5] R. B. Starr and R. J. Greenly, "Merit Rating Survey Findings," *Personnel Journal,* 17 (1939), 378–384. This article refers to a survey of 16 companies, but the tabulated results show an analysis of 18 merit-rating charts.

[6] Edwin Ewart, S. E. Seashore, and Joseph Tiffin, "A Factor Analysis of an Industrial Merit Rating Scale," *Journal of Applied Psychology,* 25 (1941), 481–486.

Table 9.2

THIRTY-FIVE RATING ITEMS USED
BY EIGHTEEN COMPANIES

A check indicates that the item appears on the company's performance appraisal chart. A number following the check indicates the weight assigned to the item. If no numbers appear it is intended to weight all items equally.

	NUMBER OF COMPANY																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
Quality		√8	√	√	√12		√50	√20				√	√	√	√10		√	√50	12
Quantity	√	√6	√	√	√10		√50	√25				√	√	√			√		11
Co-operation	√	√6	√	√			√50	√ 8	√50	√			√	√	√10				11
Initiative	√	√6	√	√	√10	√	√50	√ 6	√50	√	√								11
Dependability	√	√6	√			√		√ 9	√50	√				√	√10		√		10
Personality	√			√		√	√50		√50	√	√		√		√10	√			10
Health	√	√4	√	√	√ 4	√	√50		√50			√				√			10
Safety	√	√6	√		√ 8		√50	√ 7	√50				√	√					9
Industry		√6	√	√	√10	√	√50			√	√				√10				9
Versatility		√6	√	√		√	√50	√15					√				√		8
Leadership	√	√6	√	√	√10			√ 5	√50		√								8
Judgment	√	√4		√			√50		√50	√					√10				7
Intelligence	√	√6	√		√12					√	√					√			7
Attendance		√4	√		√ 4			√ 5					√					√10	6
Knowledge of job			√	√			√50			√	√	√							6
Potentiality	√			√						√	√			√					5
Habits		√4			√ 6	√										√		√10	5
Years of service					√			√15				√						√30	4
Loyalty	√	√4				√													3
Ability to plan	√					√			√50										3
Enthusiasm		√4				√				√									3
Trade skill		√6			√10		√50												3
Technical knowledge		√4	√	√															3
Dependents								√10				√				√			3
Punctuality		√4	√																2
General rating	√								√50										2
Tact	√					√													2
Suggestiveness	√			√															2
Knowledge of costs	√				√ 4														2
Fairness	√																		1
Knowledge of product	√																		1
Knowledge of equipment	√																		1
Knowledge of company policies	√																		1
Appearance						√													1
Place of residence												√							1
Number of rating items	21	19	15	14	13	12	12	11	10	10	7	7	7	6	6	5	4	4	
Method of rating (C = committee; S = supervisor)	S	—	C	S	S	S	C	C	C	—	S	—	S	S	S	—	S	—	
Explanatory phrases used on rating scale	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	No	No	No	

Source: *R. B. Starr and R. J. Greenly, "Merit Rating Survey Findings,"* Personnel Journal, *17 (1939), 378–384.*

Table 9.3

INTERCORRELATIONS OF THE TRAITS IN A TWELVE-ITEM MERIT-RATING SCALE

Traits	1. Safety	2. Knowledge of job	3. Versatility	4. Accuracy	5. Productivity	6. Over-all job performance	7. Industriousness	8. Initiative	9. Judgment	10. Co-operation	11. Personality	12. Health
1. Safety												
2. Knowledge of job	.61											
3. Versatility	.52	.81										
4. Accuracy	.63	.85	.80									
5. Productivity	.55	.79	.72	.81								
6. Over-all job performance	.60	.82	.80	.67	.86							
7. Industriousness	.49	.78	.71	.80	.86	.85						
8. Initiative	.54	.78	.78	.78	.80	.83	.82					
9. Judgment	.62	.80	.82	.84	.81	.88	.84	.86				
10. Co-operation	.61	.67	.68	.74	.81	.80	.80	.72	76			
11. Personality	.55	.67	.63	.70	.73	.74	.67	.72	.75	.80		
12. Health	.25	.52	.50	.84	.45	.60	.53	.77	.43	.52	.71	

Table 9.4

FACTOR LOADING AFTER ROTATION

Trait	*Factor loadings*† *Factor I*	*Factor II*	*Factor III*
1. Safety	.633	.245	−.039
2. Knowledge of job	.841	.337	−.002
3. Versatility	.795	.325	.030
4. Accuracy	.826	.448	.297
5. Productivity	.913	.039	−.042
6. Over-all job performance	.961	−.080	−.064
7. Industriousness	.906	.009	.007
8. Initiative	.887	.094	.252
9. Judgment	.897	.265	−.036
10. Co-operation	.881	−.025	.042
11. Personality	.815	.009	.259
12. Health	.573	−.016	.836

entering into Factor I reveals that over-all job performance is loaded the most heavily; productivity, second; industriousness, third; judgment, fourth; and so on down the list. If we were to attempt to name Factor I, probably the best name available would be "ability to do the present job." In other words, this factor seems to be a job-performance factor. Since the factor loadings on all of the traits entering into Factor I are high, compared with the loadings on Factors II and III (except "health" in Factor III), we infer that this factor is by far the most important of the three in determining the over-all merit rating and that, since all of the heavily loaded traits entering into Factor I deal rather specifically with ability on the job, it would seem safe to identify this factor as a job-performance factor.

Factor II contains no elements so heavily loaded as are those entering into Factor I. On the basis of the traits that have the highest loadings on Factor II (namely, safety, knowledge of job, versatility, accuracy, and judgment), it would seem that this factor deals more with quality of performance and it might be so labelled. While other names might be suggested, it is, in any event, evident that this factor reflects a different dimension of job behavior than Factor I.

Factor III was found to be significantly loaded with only one element, namely, health. For various reasons this factor can be disregarded for our present discussion. (For example, it would usually be more appropriate to determine employees' "health" by physical examination than by including it in a rating system.)

Referring back to Factors I and II, a logical conclusion one would reach is that an appraisal system in the company in question could be simplified by reducing it to two factors, namely over-all job performance and quality.

In connection with performance appraisals, Stolz[7] has emphasized that they be based entirely on the results of a man's work on the job and *never* on personality traits.

Employee Comparison Systems Where the rating scales provide for rating against some defined standard (scales on each of several traits), the employee comparison systems provide for rating employees in comparison with each other. With conventional rating scales there frequently is a tendency for the raters to pile up the ratings at one end of the scale, frequently at the higher end. To the extent that this occurs, the results are of limited value, since they do not differentiate adequately among the employees. The employee comparison systems avoid this problem completely, since the ratings of employees are relative to each other. There are three principal variations in the methods of comparing employees with each other.

---

[7] R. K. Stolz, "Can Appraisal Interviews be Made Effective?" *Personnel,* 38 (No. 2, 1961), 32–41.

Rank-order system. With this method the rater simply ranks his employees, and each employee's rating is then determined by his position in rank. If a system involving several traits is used, the rankings should be made separately for each trait. The actual process of ranking is usually facilitated by the use of small cards containing the names of men to be ranked. The cards may be arranged and rearranged by the supervisor as he ranks the men on each of the traits considered.

Paired comparison system. Some work on the application of a standard psychological method, paired comparisons, to the problem of industrial merit ratings has shown very favorable results.[8] This method ordinarily involves rating the employees on only a single general trait—over-all ability to do their present job. It may, however, be applied separately to more traits or characteristics if that is desirable. Cards or slips of paper are prepared so that each contains the names of two of the men who are to be rated. In this manner, each ratee is paired with every other one. The rater simply checks the name of the man who is considered better on the characteristic on which the men are being compared. The number of pairs of names involved when this system is used in the form described is given by:

$$\text{No. of Pairs} = \frac{N\ (N-1)}{2}$$

where N is the number of men to be rated. If 20 men are involved, the number of pairs is thus $\frac{20(20-1)}{2}$, or 190. But if 100 men are involved, the number of pairs, $\frac{100(100-1)}{2}$, or 4,950, is obviously far too great to permit expedient use of the system. Two solutions have been proposed to the problem that arises when the number of pairs becomes too unwieldy to handle. One method is to divide the original group into a number of smaller groups and set up the pairs for the smaller groups.[9] This procedure not only markedly reduces the number of pairs, but makes more convenient the rating of men in each subgroup by one or more supervisors who are much closer to the men on the job than the supervisor of 100 or more men could possibly be. Scales have been developed so that ratings of men in several subgroups, which may not be equal in size, can be converted to a common base.

A second procedure that may be used when division into smaller groups is not practicable and if there is good reason to believe that the

[8] C. H. Lawshe, Jr., N C. Kephart, and E. J. McCormick, "The Paired Comparison Technique for Rating Performance of Industrial Employees," *Journal of Applied Psychology,* 33 (1949), 69–77.

rater is sufficiently familiar with the work of each ratee on the job so that the rating will be reasonably valid, is to extract from the table of all possible pairs of ratees a patterned sample of pairs. It has been shown that the ratings obtained from such a patterned sample correlate approximately .93 with ratings obtained from the complete matrix of pairs.[10]

The paired comparison rating system was used effectively by Gruenfeld [11] in an investigation of selection tests for an executive training program. Fifty-one junior executives in training at a small midwestern college were rated independently by six instructors. The reliability of the pooled ratings of the six instructors was found to be .90.

Forced distribution system. This method has been used particularly where a department is quite large—say 40 or more employees—and where management does not deem it advisable to set up the cards for the paired comparison system. To meet this situation, the forced distribution system provides that employees be rated on one or more scales, with approximate percentages of employees stipulated for each scale location. Thus, on a scale of "performance on present job," the following percentages might be used:

Lowest 10% *Next* 20% *Middle* 40% *Next* 20% *Highest* 10%

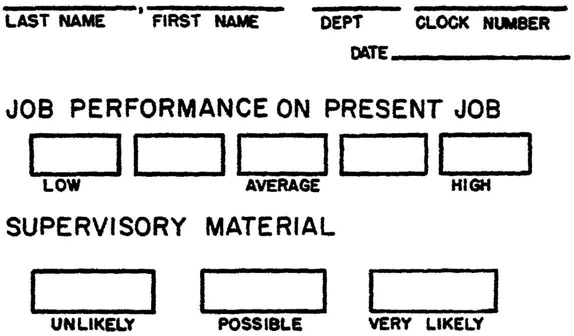
The use of a forced distribution of this type prevents some supervisors from using only the high part of the scale, others from using another part exclusively. The percentages frequently are used as guideposts rather than as rigid rules.

The forced distribution system can be used when rating a number of traits, if that is desired; in this case, the employees may be separately rated on each trait. The evidence from statistical studies such as those reported on pp. 229–232, however, is rather strongly in favor of rating employees on only a few traits. On the basis of this evidence, a simple, two-item, merit rating card has been developed. On this card, reproduced in Fig. 9.3, each employee is rated on "job performance on present job" and "supervisory possibilities." In the training program preceding the use of these cards the raters are acquainted with the percentage distribution that should be used as a guidepost in rating. It will be noted that no verbal descriptions are printed under the five boxes on the card in Fig. 9.3. Verbal descriptions are not used because, if they were used, the description under the "Low" box would necessarily have to be rather unfavorable, and many supervisors object to placing a man in the low category when it is qualified by an unfavorable comment.

---

[10] E. J. McCormick and J. A. Bachus, "Paired Comparison Ratings. I. The Effect on Ratings of Reductions in the Number of Pairs," *Journal of Applied Psychology*, 36 (1952), 123–127.

[11] L. W. Gruenfeld, "Selection of Executives for a Training Program," *Personnel Psychology*, 14 (1961), 421–431.


LAST NAME, FIRST NAME DEPT CLOCK NUMBER

DATE

JOB PERFORMANCE ON PRESENT JOB

LOW AVERAGE HIGH

SUPERVISORY MATERIAL

UNLIKELY POSSIBLE VERY LIKELY

Fig. 9.3. Form used in merit rating employees by a forced distribution system.

A second rating on "supervisory possibilities" is also called for by the card shown in Fig. 9.3. The use of forced distribution in rating this trait is not recommended, because in some departments there may be no employees who are potential supervisors, while in other departments several may have potential supervisory ability.

In using a two-characteristic rating system of this type, the two ratings should not be added together to give a single composite rating. To do so obscures the significance of the individual ratings.

Check List Rating Systems Check list rating systems embody a third basic method of rating employees. The rater is provided with descriptive statements of employee "behavior," and he is asked to indicate, in one way or another, those statements which are, or are not, descriptive of the employee in question. Thus, the rater tends to be more of a *reporter* of the *work behavior* of employees than an evaluator of their performance or their personal characteristics.

These methods are difficult and time-consuming to develop, so generally they can be justified only where they are to be used on a wide basis. They offer some advantages, however, that warrant their serious consideration for circumstances where they would be widely used.

Weighted check list. This type of merit rating system incorporates a listing of statements of work behavior such as those shown in Table 9.5. This particular check list has been prepared for rating bake shop managers. In rating an employee, the rater simply checks each statement he considers to be descriptive of the work behavior of the individual. It should be added that the scale values shown are *not* on the rating form, but are used later in deriving the employee's rating.

A brief description will be given of the procedures used in developing such a check list. First, a large number of statements are written that describe, in relatively objective terms, various aspects of work behavior,

ranging from those that are desirable to those that are undesirable. Next, these statements are "judged" by a number of "experts" on the *degree* to which they are considered to indicate favorable or unfavorable behavior. This is done by the method of "equal-appearing intervals," as originally described by Thurstone,[12] in which the judges classify the statements in categories (usually 7, 9, or 11) ranging from those they consider to be extremely favorable to those they consider to be extremely unfavorable.

Table 9.5

EXAMPLES OF ITEMS FROM WEIGHTED CHECK LIST MERIT RATING FOR BAKE SHOP MANAGER

Item	*Scale value*
His window display always has customer appeal.	8.5
He gives his employees the reasons for his decisions	6.7
Products dropped on the floor are sometimes sold.	1.4
He always gets his reports in on time.	7.9
He rarely figures the costs of his products.	1.0
He belongs to a local merchants' association.	4.9
His bakers tend to pass some of their work off onto him.	3.1
He seldom forgets what he has once been told.	7.6
He occasionally runs a selling contest among his salesgirls.	6.8
He is slow to discipline his employees even when he should.	1.9
He does not anticipate probable emergencies.	2.4
His weekly and monthly reports are sometimes inaccurate.	4.2
He should take more interest in merchandising.	3.5
He is slow at making decisions.	4.4
His bakers do not respect him.	0.8
His sales per customer are relatively high.	7.4
He has originated one or more workable new formulas.	6.4
No baking is done in his shop after 12 noon.	0.6
He encourages his employees to show initiative.	8.1
He knows how but can't teach others.	2.5
His shop is unusually neat and clean.	8.3
He often has vermin and insects in his shop.	0.8
His salesgirls sometimes fail to use wax paper to handle goods in the sales room.	3.2
He pays little attention to his customers.	1.6

Source: *Adapted from E. B. Knauft, "Construction and Use of Weighted Check-List Rating Scales for Two Industrial Situations,"* Journal of Applied Psychology, *32 (1948), 63–70.*

These judgments are then summarized and analyzed, in order to identify the statements that are most reliably judged. Table 9.6 illustrates three hypothetical examples as "judged" by ten judges. Statement A was

12 L. L. Thurstone, "Attitudes Can Be Measured," *American Journal of Sociology,* 33 (1928), 529–554.

placed in the same category (4) by all ten judges, and is therefore a highly stable item. Statement B, for which the judgments are scattered a great deal, is a very unstable item. Statement C has moderate stability.

The scale value assigned to an item usually is the mean or median of the categories in which the statement has been classified by the several judges. A possible fallacy in the use of the median for this purpose has been pointed out by Jurgensen.[13] The fallacy depends upon the fact that for some individuals many more items are checked in the top half of the scale than in the lower half. As a substitute he proposes that median values be replaced by positive and negative values obtained by subtracting the mid-value of the scale (4, 5, or 6 depending on whether there are 7, 9, or 11 categories in the scale) from each median. The merit rating score for each individual would then be the algebraic sum of these revised weights for the items checked as applying to the individual.

Table 9.6

HYPOTHETICAL EXAMPLES OF JUDGMENTS BY TEN JUDGES ON THREE CHECK LIST RATING STATEMENTS *

	Rating category								
	Unfavorable						*Favorable*		
Statement	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
A				10					
B		1	2	1	3	2	1		
C	2	5	3						

* *The number entered under each category is the number of judges who placed the statement in that category.*

In a study reported by Uhrbrock,[14] five groups of judges scaled a large number of general rating scale statements. The groups included 20 Foremen, 10 Technical Assistants, 24 University Students, and groups of 10 and of 25 University Faculty members. Correlations of about .96 and .97 among these groups indicate that the rather small groups of judges can give reliable ratings, and that judges of different backgrounds agree quite consistently regarding the "values" of rating scale statements. A few examples from the ratings by one of these groups—the foremen—are shown in Table 9.7. The "variability" column gives an index of the relative variability of the judgments (the smaller this value, the greater the degree of consistency in the judgments).

[13] C. E. Jurgensen, "A Fallacy in the Use of Median Scale Values in Employee Check Lists," *Journal of Applied Psychology*, 33 (1949), 56–58.

[14] R. S. Uhrbrock, "Standardization of 724 Rating Scale Statements," *Personnel Psychology*, 3 (1950), 285–316.

Table 9.7

SCALE VALUES OF SELECTED CHECK LIST RATING SCALE STATEMENTS AS DERIVED FROM JUDGMENTS OF TWENTY FOREMEN

Statement	*Scale * value*	*Variability †*
6. Makes the same mistakes over and over.	14	33
34. Is a clock-watcher	22	179
67. Generally looks for the easy work.	27	231
83. Can't seem to get the hang of things.	29	79
152. Is somewhat ill-natured.	36	54
171. Conduct borders on insubordination.	36	535
238. Can do good work if he (she) tries.	49	79
251. Is more interested in ideas than people.	57	541
257. Never quits ahead of time.	64	124
277. Is orderly in work habits.	69	219
297. Makes friends with others easily.	72	153
307. Can stand criticism without feeling hurt.	74	194
377. Gets production out in less than average time.	79	109
447. Submits clear, understandable reports.	84	283
498. Can concentrate under difficult conditions.	90	190
519. Is at ease in any situation.	98	179
539. Merits the very highest recommendation	107	51

* *Scale values have been multiplied by 10 to avoid decimal values.*
† *The "variability" is the square of the standard deviation of the placements multiplied by 100. It can be interpreted as a* relative *index of the consistency of judgments.*
Source: *Uhrbrock,* op. cit.

Forced-choice check list The forced-choice technique was described briefly in Chapter 7 in the discussion of personality tests. This method was originally used in employee rating, and subsequently was extended in its applications to personality tests. In forced-choice rating systems, two or more statements (typically statements of behavior) are grouped together in blocks, and the rater is asked to indicate which statement is most descriptive of the person being rated (and in some cases to indicate the least descriptive statement). In practice there are a number of variations in the method, such as differences in the number of statements in a block, and in the number and type of response required. Some variations studied by Berkshire and Highland [15] included blocks with two, three, four, and five statements, with certain groups including only favorable or unfavorable statements, but not both. An example of a block of forced-choice statements is given below. This particular illustration is from a rating form used for rating Air Force instructors; all of these statements are "favorable." [16]

[15] J. R. Berkshire and R. W. Highland, "Forced-choice Performance Rating—A Methodological Study," *Personnel Psychology,* 6 (1953), 355–378.
[16] *Ibid.*

a.) Patient with slow learners.
b.) Lectures with confidence.
c.) Keeps interest and attention of class.
d.) Acquaints classes with objective for each class in advance.

The selection of the statements for each block is based on extensive preliminary research to determine the *degree* to which each statement is considered by raters generally to be a "favorable" or "unfavorable" statement (favorability index), and the extent to which the statement, when used in a rating situation, tends to *discriminate* between above- and below-average individuals (discrimination index). The statements are then grouped together on the basis of relatively comparable favorability indexes; the statements above, for example, have been placed in the same block because they have approximately *equal favorability* indexes. The items placed in a block, however, *differ* in their *discrimination* indexes, usually only one in a block being a discriminative item.

A major advantage of this method is that a rater, attempting (consciously or unconsciously) to rate a man higher, or lower, than the man's true worth, has no way of knowing which of the statements to check to raise (or lower) the man's rating from what it should be. There are, however, certain disadvantages to the method. In the first place, since the rater does not (and should not) know how the final rating values are derived, he may resent the system as a whole, and may therefore not give it his wholehearted support. In addition, the method does not lend itself readily to counseling with the employee.

Of the various versions of this method, Berkshire and Highland[17] found that the form with four favorable statements from which the rater was to select the two most like the ratee, had advantages over other forms. It was the most resistant to bias, it yielded consistently high validities under various conditions, it had adequate reliability, and was one of the two forms best-liked by the raters.

In some situations, the forced-choice method is used in combination with other rating procedures. This has the advantage of making the combined rating procedures useful for more purposes.

Critical Incident Technique This method of employee appraisal, developed by Flanagan and Burns,[18] provides for the recording, by supervisors, of critical "behaviors" on the part of employees. Whenever an employee does something that is especially noteworthy, or especially undesirable ("critical" to either good or poor performance), a notation is made in the employee's record. These "critical behaviors" usually are classified into certain categories, such as those given in Table 9.8. When this rating system is used, supervisors note and record all "critical" in-

[17] *Ibid.*

[18] J. C. Flanagan and R. K. Burns, "The Employee Performance Record: A New Appraisal and Development Tool," *Harvard Business Review,* 33 (No. 5, 1955), 95–102.

stances of on-the-job behavior falling in any of the categories listed in Table 9.8.

Table 9.8

EXAMPLES OF CATEGORIES OF TYPES OF CRITICAL INCIDENTS

1. Physical condition.
2. Coordination.
3. Checking and inspecting.
4. Arithmetic computations
5. Learning and remembering procedures and instructions.
6. Judgment and comprehension
7. Understanding and repairing mechanical devices.
8. Improving equipment and showing inventiveness
9. Productivity.
10. Dependability.
11. Accepting supervision and organizational procedures.
12. Accuracy of reporting.
13. Response to departmental needs.
14. Getting along with others.
15. Initiative.
16. Responsibility.

Source *J. C. Flanagan and R. B. Miller, "The Performance Record. Handbook for Foremen and Supervisors of Hourly Employees" (Chicago, Ill.: Science Research Associates, 1955).*

The use of a rating method which borders on the reporting of critical incidents has been suggested by Patton [19] who points out that ratings based on personality traits such as leadership, initiative, dependability, etc., are not as satisfactory as those based on more objective measures of on-the-job performance, such as rising or falling sales, profit margins, scrap losses, employee turnover, machine down time, and the like. This is simply another way of emphasizing the importance of incidents in the appraisal process.

The use of the critical incident method in the appraisal of salesmen has been discussed by Kirchner and Dunnette.[20] Eighty-five sales managers were each asked to report as many critical incidents as possible illustrating both effective and noneffective behavior among his group of salesmen. Sixty-one usable instances were obtained. Using these instances, a rating form was prepared which the authors feel to be a very promising instrument.

Another use of critical incidents, namely the training of salesmen, has been reported by Bridgman *et al.*[21] In this investigation, critical in-

[19] Arch Patton, "How to Appraise Executive Performance," *Harvard Business Review*, **38** (No. 1, 1960), 63–70.

[20] W. K. Kirchner and M. D. Dunnette, "Using Critical Incidents to Measure Job Proficiency Factors," *Personnel*, **34** (1957), 54–59.

[21] C. S. Bridgman, J. Spaeth, P. Driscoll, and J. Fanning, "Salesmen Helped by Bringing out Jobs' Critical Incidents," *Personnel Journel*, **36** (1958), 411–414.

cidents of effective and ineffective behavior among salesmen were effectively utilized in a training program for new salesmen in the field.

Although the critical incident method does not readily lend itself to objective quantification, it offers a strong advantage for purposes of employee counseling because it provides the supervisor with a record of "specifics" to discuss with the employee.

Other Appraisal Methods There are certain other performance appraisal methods that are sometimes used.[22] One such method for example is the Field Review Method, as originally described by Wadsworth.[23] In this procedure a representative of the personnel department interviews the supervisor of each person to be evaluated, usually following an interview plan. He prepares a report on the employee based on that interview, usually sending a copy of the report to the supervisor. Habbe[24] reports what is presumably a reasonably satisfactory use of this method by a large department store.

Another method of evaluation is reported by Rowland,[25] namely, the group appraisal plan. With this plan, each man is rated by a group of supervisors working in conference. The group consists of the man's immediate supervisor and other higher managerial personnel who may have had some contact with the man on his job. Rowland feels that group appraisal tends to prevent the immediate supervisor from appraising a man without considering all of the facts that should be kept in mind. The group appraisal plan has apparently been used very successfully at the Detroit Edison Company.

SOME DANGERS OF PERFORMANCE APPRAISAL

In the previous discussion of performance appraisal there have been implications of some of their limitations or dangers. Most of these limitations are attributed to certain rating tendencies on the part of raters. These tendencies are most pronounced with conventional rating scales. In fact, some of the other rating methods have been developed primarily for the purpose of avoiding or reducing the influence of undesirable tendencies on the part of raters. The discussion of these tendencies will deal largely with conventional rating scales.

---

22 For a discussion of various methods see T. H. Whistler and Shirley F. Harper, eds., *Performance Appraisal* (New York: Holt, Rinehart, & Winston, Inc., 1962), Part 2, "Techniques of Appraisal."

23 G. W. Wadsworth, Jr., "The Field Review Method of Employee Evaluation and Internal Placement," *Personnel Journal*, 47 (1948) (Six articles in issues from June to December).

24 S. Habbe, "Merit Rating—Plus," *Management Record*, 15 (1953), 323–324.

25 V. K. Rowland, "The Mechanics of Group Appraisal," *Personnel*, 34 (No. 6, 1958), 36–43.

The Halo Effect More than 40 years ago Thorndike[26] pointed out on the basis of experimental evidence that a rater has a constant tendency to rate an individual either high or low in many traits because the rater knows (or thinks) the individual to be high or low in some specific or particular trait. Thorndike called this tendency the "halo" effect. Applied to the industrial situation, Thorndike's statement means that if the supervisor regards an employee as very satisfactory in terms of a trait such as general personality or tact, he is likely to rate the employee high also in such traits as productivity, ingenuity, inventiveness, adaptability, and perhaps many other traits. In other words, it is difficult for any rater—particularly an untrained rater—to isolate and rate separately the various traits that an employee may possess. The use of a merit-rating chart is likely to increase the ability of a foreman to make an analytical judgment; but we know that even under the most favorable conditions the halo effect will be present to some extent and that its results will be most prominent where ratings have been made by those unfamiliar with its very existence.

The halo effect can be minimized in several ways. If a rating scale system is used, it is sometimes advisable to have each supervisor rate all his men on one trait before going on to the second trait, on the second trait before going on to the third, and so on. Since this method causes the supervisor to think of all his men in connection with a given trait alone rather than to think of each man as a whole, the effect of this general change in point of view is a reduction of the halo effect. The halo effect can also be minimized by arranging the chart itself so that the desirable end of some traits is on the right-hand side of the scale whereas the desirable end of other traits is on the left-hand side of the scale. This procedure prevents a supervisor from checking down a column on the right-hand side for a generally desirable employee, or down the left-hand side for a generally undesirable employee.

However, if this format is used, an introductory statement should call the rater's attention to the fact that the same end of the scale is not always "good." Guilford[27] feels that varying the "good" and "bad" ends of the scale has not been shown experimentally to reduce the halo effect and that even experienced raters are often confused by such an arrangement. Such confusion can be eliminated, however, if a carefully worded explanatory statement appears at the top of the chart.

The operation of the halo effect in an actual set of ratings is shown in Fig. 9.4. This figure reveals graphically, for 18 men selected at random from a large industrial organization, the relationship between ratings on

[26] E. L. Thorndike, "A Constant Error in Psychological Ratings," *Journal of Applied Psychology*, 4 (1920), 25–29.

[27] J. P. Guilford, *Psychometric Methods*, 2nd ed. (New York: McGraw-Hill Book Company, 1954), pp. 267–268.

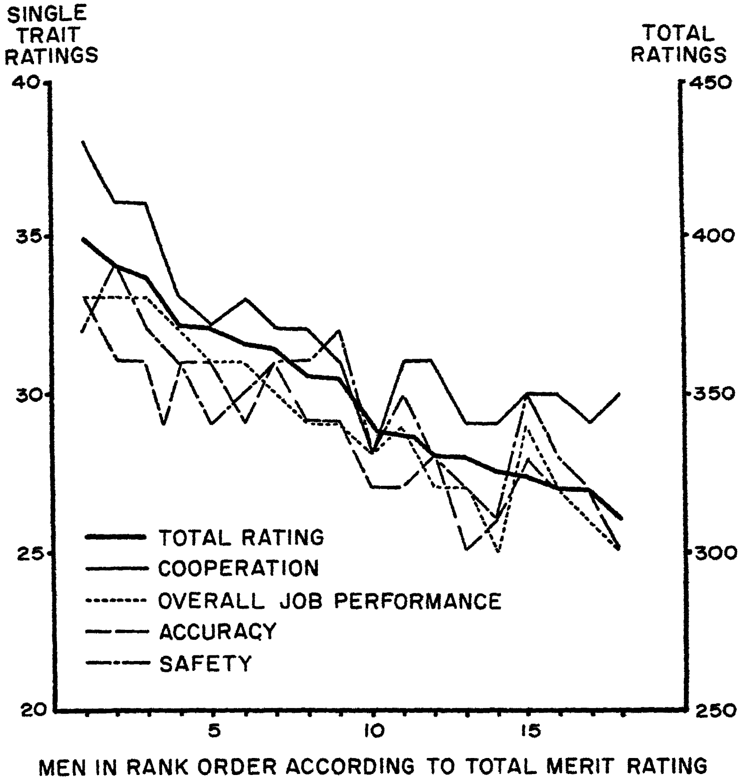


Fig. 9.4. Operation of the "halo effect" in industrial merit rating.

over-all job performance, accuracy, safety, and co-operation, on the one hand, and on the other, an over-all merit rating that included the four traits mentioned above along with eight others that need not be identified. The 18 men are arranged in rank order from left to right according to their over-all merit rating. These ratings are indicated by the heavy black line. The ratings of the 18 men on the four traits are indicated by the remaining lines. It will be seen from an inspection of Fig. 9.4 that those individuals who were rated high in any given trait, say co-operation, were also rated about equally high on all of the other traits shown. Of course, it may theoretically be possible that the individual high in one trait is actually high in all of the other traits, but it seems much more likely that the halo effect, rather than any real relationship among the traits, is operating here.

In general the halo effect results in ratings on various traits or attributes having higher correlations with each other than would other-

wise be the case. To some degree this tendency probably can be minimized somewhat by training the raters; for example, a study by Driver[28] of a set of ratings obtained by raters who had been given seven hours of intensive training in rating methods revealed much lower intercorrelations than those shown in Table 9.3.

The general conclusion with regard to the halo effect is that unless supervisors have been trained very carefully they may allow their general impression of an employee, or their evaluation of one of his traits, to influence their ratings on other traits as well.

Constant Error Some raters tend to be lenient in making ratings, thereby giving many employees "high" ratings. Other raters tend to be "tough," and give primarily "low" ratings. Unless a suitable correction is made for such tendencies of different raters, ratings of different men that have been made by different raters cannot reasonably be compared. For example, in using a 12-factor rating system, each factor having a maximum value of 50 points, the ratings made by one supervisor averaged 405 points, while the ratings made by another supervisor (on a presumably comparable group of men) averaged 295 points. The ratings by these two supervisors are not comparable unless they are converted to a common base.

Influence of Other Factors on Ratings It sometimes happens that raters are influenced, unknowingly, by various factors, in some cases by factors that actually should not have such influence. Among such factors are department, job, age, and length of experience.

Departmental differences. Very frequently the ratings turned in from different departments in a given plant differ markedly from one department to another. This discrepancy may be due in part to actual differences in the merit of employees in the various departments, but it may also be due in part to differences in standards or interpretation of the rating scale among the departments. Whatever may be the cause, when such differences do occur it is usually desirable to evaluate a given employee's rating in terms of the other ratings from the department in which he is working rather than in terms of the ratings obtained from the plant as a whole. This situation is graphically illustrated in Fig. 9.5, which shows the distributions of total performance ratings obtained from three departments of a plant made up of 14 departments and employing approximately 10,000 men. The three departments selected for illustration are engineering, maintenance, and plant protection. These three have been selected to show the marked differences that may be found in ratings from one department to another. The difficulty of interpreting the significance of performance rating without reference to the department from which it was obtained may be readily seen from inspecting Fig. 9.5. For

---

[28] Unpublished study by R. S. Driver, Atlantic Refining Co., Philadelphia.

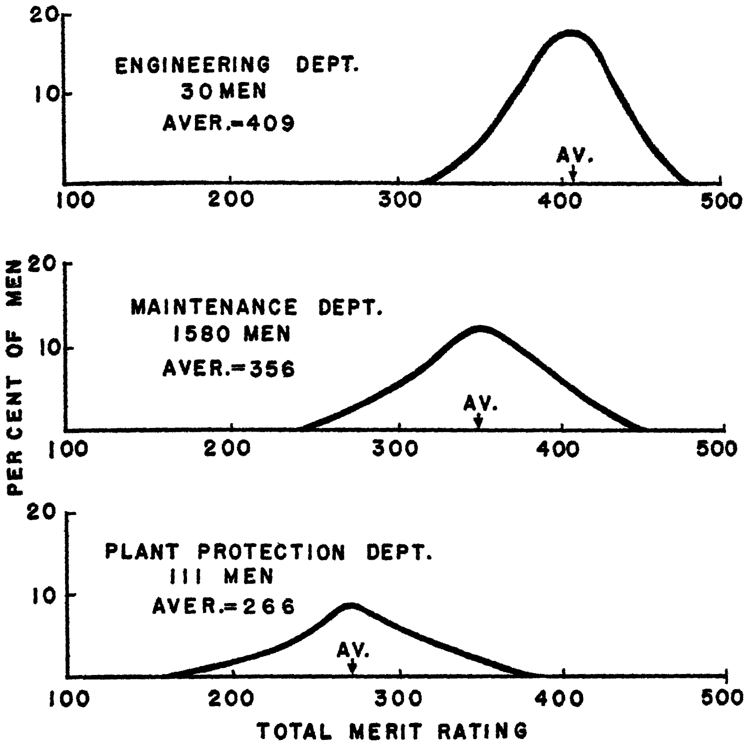


Fig. 9.5. Differences in merit ratings among departments in a steel mill.

example, a rating of 350 would be a very low rating for a man in the engineering department. The same rating would be approximately average for a man in the maintenance department, and it would be very high for a man in the plant-protection department. Since one major function of performance appraisals is to indicate how well—in relation to other employees—an employee is doing his present job, a fair and reasonable basis for comparing ratings of different men must be employed. When ratings differ markedly from one department to another, evaluation of any rating should be in terms of the department from which it was obtained.

Job differences. Another source of possible difficulty closely related to the matter of departmental differences is the variation in rating often found from one job to another. When employees on any given job are consistently given higher merit ratings than are employees on other jobs, such job differences should be considered in evaluating the rating of any given employee. Fig. 9.6 illustrates this situation for 51 jobs in a sheet

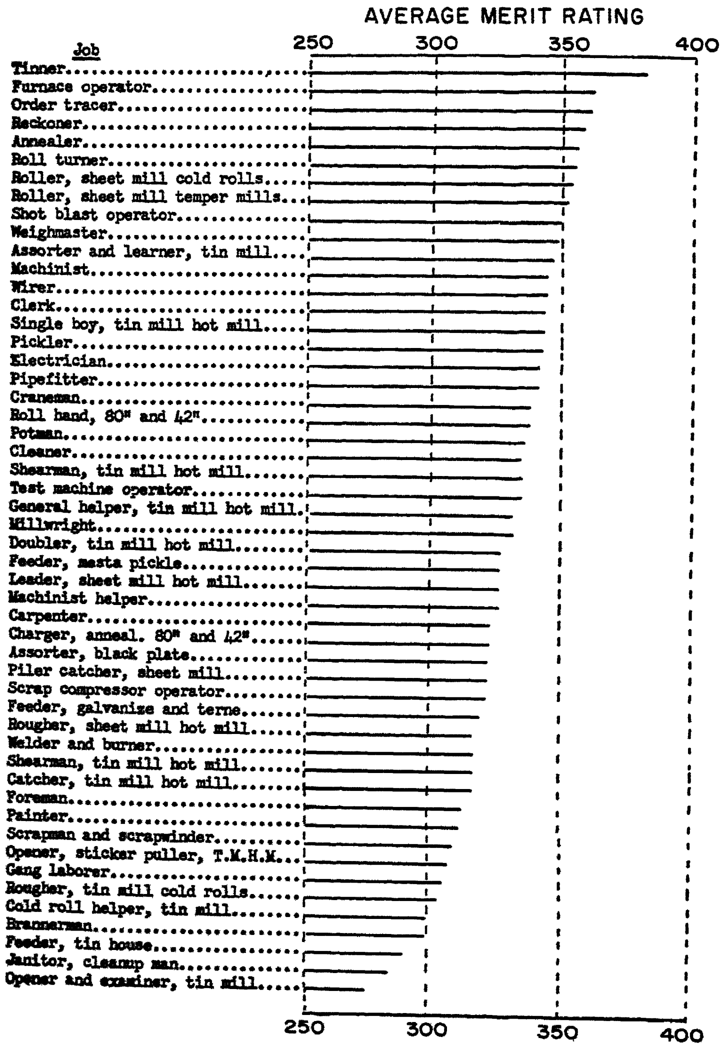


Fig. 9.6. Differences in average merit rating of employees on 51 jobs in a sheet and tin mill.

and tin mill. The 51 jobs studied are arranged in order from the one receiving the highest average rating (tinner) at the top to the one receiving the lowest average rating (opener and examiner) at the bottom. It will be noted that there is a variation from 275 to 385, or 110 points, in average performance rating from the lowest to the highest average rating. From these differences it is clear that a rating of 300 is very high for an employee who is on one of the jobs located near the bottom of the list, but that 300 is very low for an employee on one of the jobs near the top of the list. The implication of this fact is that the performance rating of an employee should be evaluated in relation to the ratings of other employees on the same job or on jobs where the employees are given approximately the same average rating.

Age of employees. The age of an employee is another factor that is often related to the rating he receives. In one set of approximately 9,000 merit ratings obtained from a single industrial plant, the relationship between total rating and age is graphed (Fig. 9.7). This figure shows that

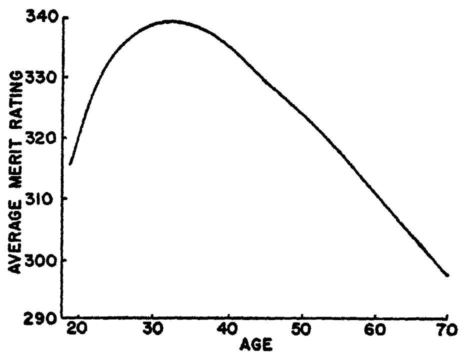


Fig. 9.7. Relation between age and average merit rating for 9,000 steel workers.

employees between the ages of 30 and 35, other things being equal, have higher ratings than those who are either younger or older. The possible causative factors are of course not evident from the data as such. While one might infer that people in their early thirties might perform generally better than younger or older persons, this is not necessarily a valid inference. For example, it is possible that the generally more able individuals of higher ages might have been promoted into other positions that were not included in the rating group, thus leaving those of lower general abilities. Regardless of the possible explanation, however, the point is made here that where a relationship between age and ratings such as this exists, it should be known, in order that it can be taken

into account in the use of the ratings, such as by comparing the ratings among those persons of a given age group.

Other factors. The preceding factors that should be considered in evaluating the performance rating of employees have been mentioned as illustrative of the kind of factors that have been found in a number of investigations to be related to such ratings. Still other factors exist that may affect merit ratings in any given plant. Fig. 9.8, for example, illus-

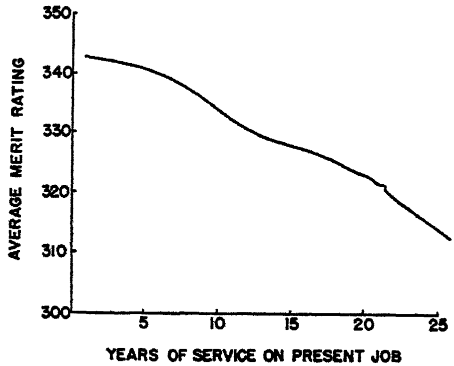


Fig. 9.8. Relation between years of service on present job and average merit rating for 9,000 steel workers.

trates the relationship found in one plant between ratings and length of service on the present job. In definite contradiction to the opinion of management before these results were obtained, this chart shows a progressive lowering of ratings as the length of service on the job continues. In other words, the longer a given employee remains on a certain job, the lower the merit rating he is likely to receive from his supervisor. The probable explanation for this relationship is that only those employees who are unqualified for promotion or transfer to another or more important job are likely to remain on their present jobs for a long period of time. Or perhaps, in this particular plant, the employees with longer service on their present jobs are in general older employees and may be unable to do the work as effectively as younger men. Whatever may be the cause of the relationship, it is clear that it exists and that it should, therefore, be considered in evaluating a given employee's rating. A rating, such as 335, that might be well below average for an employee of short service on the job, would be well above average for an employee who has been on the job fifteen or twenty years.

Fig. 9.9 graphs the relationship between rating and total service in the plant. It is clear that the fluctuation in average merit rating with total plant service is much less marked in the case of this relationship

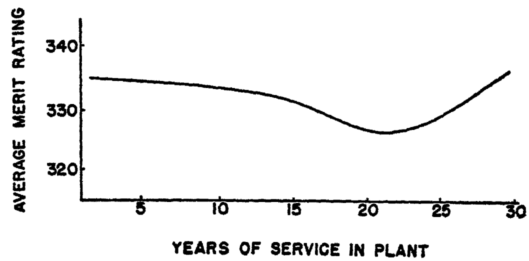


Fig. 9.9. Relation between years of plant service and average merit rating for 9,000 steel workers.

than in the case of the relationship with length of service on the present job. There was no marked or systematic change in total rating with total plant service.

It is not meant to imply from the results summarized in the preceding sections that the relationships found in the studies reported exist in exactly this same form in all industries. Indeed, there is every reason to expect that these relationships are not universal and that opposite trends may occur in many plants. The point to keep in mind is not the universality of the particular trends discussed but the fact that no company can properly interpret the results of the merit ratings in its own plants without definite knowledge of what trends and extraneous factors are related to these ratings in that particular organization.

ADJUSTING RATINGS FOR UNDESIRED INFLUENCES

We can see from the above discussion that ratings may be influenced by the rater's "constant" error, or may be related to such factors as the job, the department, and age. In order to determine whether there *are* differences in the ratings of employees by rater, job, and department, it is necessary to make an analysis of the ratings such as those previously shown in Figs. 9.5, 9.6, 9.7, 9.8, and 9.9.

Where it is found that ratings *do* differ significantly for the employees in different categories (such as employees rated by different raters, those in different departments, those on different jobs, and those of different ages), there are two possible explanations for the differences. In the first place it is possible that the differences are real differences—that the employees in one category in general *are* better than those in another category. The other possible explanation is that the ratings have been influenced by spurious factors such as raters' tendencies, and that the differences are simply functions of these influences; in this case the systematic differences—say, in the ratings given by two raters—are not true differences. Where there are significant differences in ratings of

employees in various rating groups (such as employees rated by two or more raters), the fact of these differences is immaterial *if* the ratings are to be used *entirely within the context of the group.* If the ratings are to be *compared between or among* rating groups, however, it is pertinent to make an informal judgment as to whether there is any justification for believing that the differences are *real* or *spurious.* In the absence of evidence to the contrary, it is probably most justifiable to assume that they are attributable to raters' tendencies or similar factors that should be ruled out.

In such a case, the ratings should not be compared across rating groups *unless* they are adjusted for the differences. There are at least two ways of doing this.

Adjusting for Differences in Means Let us continue to use, as an example, the situation in which the ratings given by different raters differ significantly. In such a case, one solution is to determine the average of the ratings given by each rater, and compute the difference between *each rater's* average, and the average of *all* raters. This difference can then be added to or subtracted from the ratings given by a particular rater, in order to bring his ratings into alignment with those of other raters. This simple adjustment is satisfactory if the *variabilities* of the ratings given by the different raters are about the same.[29]

Adjusting for Differences by Standard Scores A more systematic method of adjusting for such differences consists of converting all ratings to a common set of numerical values. Some type of *standard score* (comparable score) may be used for this purpose. There are various types of standard scores, such as *z*-scores. Standard scores, including *z*-scores, indicate *relatively* the position of individual cases in a distribution. Such scores are based on deviations of individual cases from the mean *in standard deviation units.* A "standard deviation" is a statistical index of the degree of variability of the cases within a distribution.[30] It is expressed in terms of the numerical values of the original distribution. In a relatively normal distribution, two-thirds of the cases fall within one standard deviation above and below the mean, about 95 per cent are within two standard deviations above and below the mean, and about 99 per cent fall within three standard deviations. Thus, regardless of what the mean of a distribution is, or what the magnitude of its standard deviation, it is possible to express the deviation of any given numerical value in terms of the number of standard deviation units it is above or below the mean. A *z*-score is simply the deviation of a given raw score from the mean expressed in terms of standard deviation units.

---

[29] The variabilities in the ratings can be compared by comparing their *standard deviations.* See Appendix A, pp. 627–629.

[30] *Ibid.*

Let us now see how this helps us in comparing the merit ratings of employees as rated by a "tough" rater and as rated by a "lenient" rater. Let us suppose that the distributions A and B of Fig. 9.10 represent,

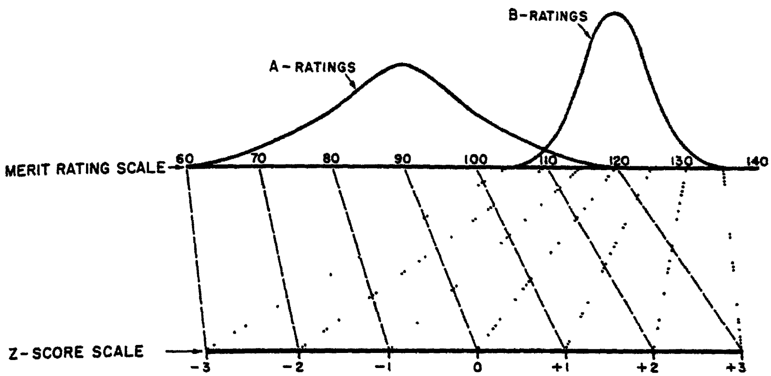


Fig. 9.10. Illustration of conversion of two sets of ratings (A and B) to a common scale of z-scores

respectively, the total ratings given to their respective groups by rater A and rater B. We can see clearly that a rating of, let us say, 110 by itself, is meaningless unless we relate it to the distribution of which it is a part (it means a very high rating by rater A, and a very low rating by rater B). These two distributions, however, can both be converted to *z*-scores, as illustrated at certain points on the distributions by the broken lines and dotted lines from the merit rating scale to the *z*-score scale below it. We can now see that a rating of 110 by rater A would mean a *z*-score of *plus* 2, but by rater B would mean a *z*-score of *minus* 2, and that a rating of 100 from rater A would correspond to a rating of 125 from rater B, since both convert to *z*-scores of plus 1.

Similar conversions can be made for ratings of people from different rating groups, such as for different jobs or departments.

THE "WEIGHTING" OF TRAITS

In the conventional rating scales for performance appraisal purposes it is customary to rate employees on each of several traits or characteristics. Usually it is considered desirable to weight each of these traits according to what seems to be the relative importance of each for success of employees in that particular organization. Thus one company might weight initiative 20 points and judgment 5 points, whereas an-

other organization might reverse this weighting of items. Table 9.2 shows that of the eighteen charts analyzed, four, or 22 per cent, incorporated some sort of differential weighting for the various items. The remaining charts use either equal weightings or no weightings at all. Either of these methods is presumably intended to give the same importance to the several traits included on the chart.

At least two difficulties arise in connection with weighting items in a performance appraisal system. The first of these is that items are not necessarily weighted equally when all are given the same maximum numerical value, nor are they necessarily weighted in the way intended when a predetermined set of maximum values for each is used. In combining scores—whether they are performance rating items, test scores, production records, or any other set of values—the scores weight themselves automatically in proportion to their respective variability. Expressed in statistical terms, the scores weight themselves in proportion to their respective standard deviations. Therefore, if the variability of all employees on one rating item, say, judgment, is twice as large as the corresponding variability of all employees on some other item, say, initiative, a direct combination of ratings for any employee on these two traits is actually weighting the judgment ratings twice as heavily as the initiative ratings. It might, of course, be the wish of management to weight those two items in this proportion, but it is unlikely that the chance weights that creep into a set of ratings as a function of their respective variabilities will weight the various traits in the manner desired by management.

The statistical reason for weights being determined by variability of the ratings is discussed in some detail in Appendix A, p. 629. An example here may further clarify the principle.

Suppose 1,000 men have been rated on two traits, namely, judgment and initiative. Each man has been rated on each trait on a 50-point scale. Suppose, for the present illustration, that all of the men have received ratings on initiative of between 30 and 35 points. Suppose, further, that the judgment ratings vary from 25 to 45. If we now combine for each man his rating on initiative and his rating on judgment we will obtain a combination rating in which it has often been assumed (because each trait was originally rated on a 50-point scale) that the two ratings are weighted equally. Under these circumstances, however, the traits are not weighted equally at all. The judgment ratings, which vary over a range of 20 points—from 25 to 45—will have approximately four times as much effect on the total rating as the ratings on initiative, which vary over a range of only 5 points—from 30 to 35. A method of combining the ratings so that they may be weighted equally is described in Appendix A. If this method seems rather complicated, it can only be said that the rating of human characteristics and performance is complicated and is

difficult enough to accomplish satisfactorily even when all statistical safeguards are followed.

The fact that unknown weights for the various items in a performance appraisal program not only *may* but *do* creep in, if not guarded against statistically, is proved by an analysis of ratings of 1,000 employees in a steel mill. The variability of the ratings on each of twelve traits was computed and the standard deviations are shown in Table 9.9.

Table 9.9

STANDARD DEVIATIONS OF RATINGS OF EMPLOYEES IN TWELVE TRAITS

Trait	*Standard deviation*	*Relative weight*
1. Safety	2 24	1.00
2. Knowledge of job	2 77	1.24
3. Versatility	2.88	1.29
4. Accuracy	2 69	1.20
5. Productivity	2.58	1 15
6. Over-all job performance	2 63	1.18
7. Industriousness	2 96	1.32
8. Initiative	3.08	1 38
9. Judgment	2 68	1 20
10. Co-operation	2.72	1.22
11. Personality	2.51	1 12
12. Health	3 14	1.40

The employees had been rated on each of these traits on a 50-point scale, and it had been assumed by the management that this procedure resulted in total ratings that were influenced in an equal amount by each trait. Actually, the health ratings, which had the largest standard deviation, were exercising 40 per cent more effect on the total ratings than the safety ratings, which had the smallest standard deviation. Next in importance were the ratings on initiative, which were exercising 38 per cent more effect than the safety ratings. The relative weights actually exercised by each of the 12 factors are given in the final column of Table 9.9. It is doubtful whether the management of this company would have agreed upon this weighting of items if the matter had been discussed when the chart was constructed.

The simple adding of ratings for several factors not only fails to weight the traits equally (except occasionally and by chance) but also fails to give them any preassigned weights that might have been decided upon and crystallized in the form of a maximum value that each trait may receive. Suppose, for example, that management has decided that *accuracy* is twice as important as *production* and therefore has adopted a rating system in which *accuracy* is rated on a 40-point scale and *production* on a 20-point scale. This arrangement will not necessarily result in

accuracy being weighted twice as heavily as *production, for the relative weights of the traits are determined by the variability or spread of each and not by the maximum values assigned to each.* It would be quite possible in the situation described above for the production ratings to assume the heavier weights if the original ratings were directly added.

A second hazard related to the matter of weighting items in a performance appraisal system, even when proper steps have been taken to insure the functional operation of the weights decided upon, deals with the weighting of factors for employees on different jobs. For example, the qualities that contribute to performance of, say, a private secretary of a plant manager, as opposed to those required for a receptionist, would argue for the evaluation of their individual performance on the basis of different rating factors. Where different personality characteristics or work behaviors are relevant to success on different types of jobs, the performance appraisal system preferably should take such differences into account by keying the system to the job requirements. Otherwise, in adding up ratings on different factors to derive a total rating, one might unknowingly give excessive weight to some factors and too little weight to others. This problem is, of course, largely eliminated where the performance appraisal system provides simply for rating on over-all job performance. In such a case, the evaluator himself can take into account the personal characteristics and job behaviors that he considers relevant to the particular position. The halo effect, of course, tends to result in such a generalized, over-all, evaluation. In fact, it was suggested by Bingham [31] a number of years ago that the halo effect may be the saving grace of an appraisal system. In other words, when a supervisor begins to rate any given employee on several traits, he may, and probably does, first of all center his attention upon the particular trait or traits that are necessary for the employee on his particular job. After these traits have been rated, all of the remaining traits to be rated automatically fall in line in terms of where the key traits have been rated. But the old axiom that one cannot eat his cake and have it too prevails in the use of an itemized *vs.* an over-all performance appraisal system. To the extent that the system really does fractionate an employee's characteristics, it is unlikely to give an accurate indication of his ability on his present job; whereas to the extent that the system fails to fractionate the traits, owing to the halo effect, it is likely to work well in terms of giving an adequate over-all indication. Some companies have solved this problem by using a dual system—that is, one in which the employee is first rated on over-all ability on his present job before ratings of a fractionated sort are obtained. Such a dual system will accomplish both results with little

[31] W. V. Bingham, "Halo, Invalid and Valid," *Journal of Applied Psychology*, 23 (1939), 221–228.

more labor on the part of supervisors than is included in any of the systems now in use.

RELIABILITY AND VALIDITY OF PERFORMANCE APPRAISALS

The concepts of *reliability* and *validity*, as previously discussed in connection with tests, have corresponding meanings in relationship to performance appraisals. The reliability of ratings is the consistency with which the ratings are made, either by different raters, or by one rater at different times. The validity of ratings is the degree to which they are truly indicative of the intrinsic "merit" of employees.

Reliability of Ratings It can readily be understood that ratings must have an adequate degree of reliability in order to be useful for their intended purposes. For example, if on one occasion the rating for an individual is high and on another occasion it is low—when the individual himself has not changed—one cannot then accept any given rating with confidence. The implication of this statement is that the more unreliable the ratings are, the greater must be the differences in ratings (such as between different employees, or for the same employee at different times) before we are justified in assuming that the differences in ratings reflect *true* differences in merit.

The reliability of ratings is to some degree a function of the type of rating method used. For example, the reliability of rating scales is typically somewhat lower than for employee comparison systems. Data on the reliability of ratings on each item of a twelve-item rating chart, as well as the reliability of the total rating, are given in Table 9.10. These reliability coefficients shown in Table 9.10 were obtained by correlating results from pairs of raters who had rated the same employees. The data are based on a total of 92 raters and 4,500 rated employees.

Somewhat higher reliabilities typically are obtained with employee comparison systems. With a ranking procedure, for example, ratings with reliabilities as high as .85 to .95 have been obtained.[32] Experience with the paired comparison method has indicated that it frequently results in relatively acceptable reliability. In one study, for example,[33] eight men were rated in common by supervisors A and B, eight others were rated in common by supervisors A and C, and eight others were rated in common by supervisors B and C. The results of this study are summarized in

[32] Personal communication from Dr. H. C. Taylor of the Upjohn Foundation for Community Research.

[33] C. H. Lawshe, Jr., N. C. Kephart, and E. J. McCormick, "The Paired Comparison Technique for Rating Performance of Industrial Employees," *Journal of Applied Psychology*, 33 (1949), 69–77.

Table 9.10

RELIABILITY OF EACH ITEM OF A 12-ITEM EMPLOYEE-RATING SCALE, AND TOTAL RATING BASED ON THE SUM OF ALL TWELVE ITEMS

Trait	*Reliability*
1. Safety	.35
2. Knowledge of job	.46
3. Versatility	.47
4. Accuracy	.45
5. Productivity	.46
6. Over-all job performance	.46
7. Industriousness	.47
8. Initiative	48
9. Judgment	.45
10. Co-operation	.37
11. Personality	39
12. Health	.36
Total	.55

Table 9.11. The average correlation of .83 obtained in this study is typical of the reliability that has been achieved when this rating system is used.

Table 9.11

RANK-ORDER CORRELATIONS BETWEEN PAIRED COMPARISON RATINGS OF EIGHT MEN BY TWO DIFFERENT SUPERVISORS

Raters	*No. of men*	*Rank order correlation* *
A and B	8	.81
A and C	8	.83
B and C	8	.86
Average		.83

* *See Appendix A, p. 637, for a description of the rank-order coefficient of correlation.*

All of these facts considered together point up the importance of knowing the reliability of ratings obtained with a performance appraisal system. This statement is not intended to imply that performance ratings are of no value, but it does mean that one should be aware of their reliability and should not attempt to use such ratings in a way that presupposes a higher reliability than they actually possess.

In proportion as ratings are unreliable, it is not a valid procedure to consider a slight change in rating from one time to another as indicating a real change in the merit of the employee. By means of a simple statistical procedure, it is possible to obtain for performance ratings what is known

as the "standard error of measurement." Unless an employee's rating changes by at least three standard errors of measurement from one rating to another, it is unsafe to assume that any real change has occurred.

Validity of Ratings Quantitative evidence of the validity of ratings usually is difficult, if not impossible, to obtain. In fact, merit ratings sometimes are used *because* there are no other criteria of employee merit available. Under such circumstances the *validity* of merit ratings may have to be *inferred* from their *reliability*. By this we simply mean that if there is a high degree of consistency between and among raters in rating the same employees, it must be assumed that, *because* of this high reliability, the ratings presumably are indicative of true relative merit (validity). Although high reliability may be used as a basis for inferring the validity of ratings, it does not necessarily follow that low inter-rater reliability indicates that all of the ratings are of low validity. It is possible that low correlation between and among raters may simply be the result of the fact that some of the raters do not know the employees well enough to rate them adequately.

Combining Ratings from Different Raters It is sometimes the practice to have two or more raters rate each employee, and then combine, or pool, their ratings to obtain a single, pooled rating for each employee. There is evidence to support the contention that pooled ratings *by competent raters* are better than single ratings. This was indicated, for example, in a study by Bayroff, Haggerty, and Rundquist.[34] In this investigation 400 Army officers (primarily majors and lieutenant colonels, attending a Command and General Staff College) were rated experimentally by various methods. They were also ranked by about 20 of their fellow-officers on "value to the Army." These rankings served as a criterion for "validating" the ratings that were made separately. The validity of an eight-step rating scale was determined by correlating ratings from this scale with the criterion rankings. The average correlation for individual raters was about .52 (.52 for one group of 200, and .53 for another group of 200). When combining the ratings of ten raters, however, the correlations were .89 and .84. Thus, if there are several *competent* raters who can rate each employee (as there were in this case) it would seem worth the time and trouble to obtain such multiple ratings.

There is, however, a big "if" in this statement, for it frequently is the case in industry that the only person who really can properly rate an employee is his immediate supervisor. In fact, the very organization of a modern industrial establishment is such that the more ratings one attempts to obtain on a given employee, the further away from that employee in terms of contact with him one must go in order to obtain ad-

[34] A. G. Bayroff, Helen R. Haggerty, and E. A. Rundquist, "Validity of Ratings as Related to Rating Techniques and Conditions," *Personnel Psychology*, 7 (1954), 93–113.

ditional raters. In this connection, Whitla and Tirrell[35] report a study in which it was found that the raters closest to the ratees (in terms of organizational level) were able to rate the ratees better than those raters who were at higher levels of the organization. While the pooling of ratings generally increases their validity (if the raters are reasonably familiar with the ratees), there is considerable danger in assuming that pooling will automatically increase the validity of employee ratings. The raters need to be those who are closely familiar with the ratees and their work performance. Usually such persons will be the immediate supervisors of the ratees. Some performance appraisal systems provide for the rater to indicate how well he knows the ratee.

A PERFORMANCE-RATING CASE STUDY

The performance-rating program of the Acme Steel Company, which has been in operation for several years, includes safeguards against most of the extraneous influences discussed in this chapter. A rating system which includes ten factors is used. The factors are:

1. Accuracy of work.
2. Quantity of work.
3. Use of working time.
4. Ability to work with others.
5. Ability to learn.
6. Safety.
7. Initiative and acceptance of responsibility.
8. Conduct on the job.
9. Care and use of equipment, tools, material, supplies, power.
10. Attendance and punctuality.

Each man is rated independently by two or more supervisors who are familiar with the man's work. Each rating is then converted to a common base to eliminate the raters' tendencies. The average corrected ratings by the two or more raters for each man are then determined. The men on a given job or group of jobs in the standard promotional sequence are then listed in order of their corrected average rating. The men in approximately the highest 10 per cent on this list are considered to be equal from the standpoint of the ratings, and the man in this group with the longest applicable seniority is given the option of accepting any promotion that may be made in the particular promotional sequence involved.

SUGGESTIONS FOR IMPROVING A PERFORMANCE-RATING PROGRAM

Many of the facts discussed in the preceding pages of this chapter may be put to practical use by deriving from them a number of sugges-

[35] D. K. Whitla and J. E. Tirrell, "The Validity of Ratings of Several Levels of Supervisors," *Personnel Psychology*, 6 (1954), 461–466.

tions for installing and operating a merit-rating system. Some of the more important suggestions are summarized below:

1. Train the raters. Rating people is a skill, and one that is not easy to learn. Good production men and even good supervisors are not necessarily good raters *unless they have been taught how to rate.* Even at best, rating is subjective and personal. Prejudice and bias can never be completely removed, and in the absence of training they may completely distort many ratings. Companies that have had all raters attend a systematic training program of from six to eight hours before making any ratings have found, without exception, that the time and money spent on this training was a good investment.

A training program for raters preferably should cover such subjects as:

(a) The nature of the halo effect and of constant errors, and how to guard against them.
(b) The desirability of placing emphasis on observable behaviors where possible, as opposed to emphasis on personal traits of the ratees.
(c) The importance of spreading the ratings over the whole range of ratings—but doing so discriminantly.

2. In addition to training as such, another procedure that increases the accuracy of ratings is to have them made in conference or under supervision. Adoption of this system does not mean that supervisors confer with each other about their men while the ratings are being made, but only that the ratings are made while the raters are gathered for this purpose under the guidance of someone thoroughly familiar with the system. By the use of this procedure, several difficulties will be avoided. Careful thought will be given to the problem, and the hasty checking that often takes place if the supervisor is expected to "find the time" to make the ratings will be eliminated.

3. After the ratings have been made, be extremely cautious in comparing the ratings of men in different departments or on different jobs. The need for this caution is brought out on p. 244. Unless a correction has been made for a man's job, it is unfair to compare his rating with ratings of men on different jobs.

4. Interpret the ratings in terms of actual job demands. A man should not be penalized because a rating shows him to be weak in some ability that his job does not require.

5. Avoid the use of numerical values in discussing ratings with employees. Numerical values give to ratings an appearance of greater accuracy than they in reality possess. The reason for this caution is discussed on p. 264.

6. It is ordinarily wise to omit from the chart factors concerning which objective information can be obtained without using ratings. If

production records are available, it is better to use these in evaluating the production of an employee than to rely on a supervisor's judgment. The same thing is true of quality of work, if an actual record of rejects or waste material is available, and of health.

7. Avoid pooling poor ratings with good ones. The use of *one* satisfactory rating is better than the pooling of many, if some of those pooled are likely to be in error. This subject is discussed on p. 257.

8. A final suggestion relative to the most effective use of merit ratings based on rating scales, or indeed of merit ratings based on any system, is to make use of all possible supplementary information to be obtained from other sources. A very great deal of information about the performance of employees often remains unassembled and unused in many plants. Information on accidents and hospital visits, absenteeism, training time and costs, production, quality, and so forth, frequently is available, and can be very helpful in interpreting or supplementing merit ratings. Job-information tests are also being found useful by a number of plants as one of the tools to be used in upgrading and promoting. In some companies the union has endorsed the use of tests for this purpose when performance ratings alone were not accepted. In a grievance hearing, an unsupported rating presented to an arbitrator as reason for an alleged discrimination against an employee with seniority is usually very difficult to justify. But a rating supported by one or more test scores on job-information tests or by records covering production, quality, and so forth, often forms a persuasive and conclusive case.

THE USES OF PERFORMANCE APPRAISALS

For a performance appraisal program to serve its purposes of helping employees to develop in their present jobs and perhaps to be able to accept more responsible positions, it is necessary to feed back to them relevant information about their appraisals. Sound psychological principles of learning indicate that, for improvement (learning) to take place, knowledge about one's previous performance is essential. Performance appraisals can serve this purpose.

The Effects of Feedback of Appraisal Information The effects of feedback of appraisal information have been pointed up by the results of various studies. One such study is reported by Moon and Hariton.[36] (This study actually deals with the effect of a training program set up to equip supervisors with the understanding and skills needed to "feed back" appraisals in a manner designed to stimulate growth and development in the men.)

Fifty managers (supervisors) were trained in groups with about 12 in each group. The training involved 30 hours of instruction spread over 2

[36] G. G. Moon and T. Hariton, "Evaluating an Appraisal and Feedback Training Program," *Personnel*, 35 (No. 3, 1958), 34–41.

weeks of half-day sessions. The sessions covered how to guide the man being appraised in formulating plans for self-improvement. Ways of helping him to achieve greater personal and on-the-job efficiency were studied and included day-to-day coaching, job enlargement and rotation, appropriate training courses, and participation in professional and civic organizations.

The program included not only essential *principles* but also provided practice in developing the necessary interviewing *skills*. Each participant practiced the recommended procedures in realistic role-playing sessions under close observation of his colleagues and trainers.

Two years after the above training program was inaugurated, an evaluation of it was attempted from two points of view. It was felt that the opinions of the appraised men about changes in their supervisors' attitudes and behavior would be a better measure than what the supervisors themselves thought about the program. Thus the core of the evaluation of the program was a questionnaire designed to obtain the employees' views about changes in their supervisors' attitudes and behavior after the program was installed. However, it was also felt that the opinions of the supervisors themselves would add to the picture. Accordingly, an additional questionnaire survey was conducted among the supervisors.

The latter survey, which might be summarized first, showed that 97 per cent of the supervisors felt that the training made it easier for them to appraise the performance of their employees, 75 per cent said the training made it easier for them to talk to employees about job problems, and 72 per cent said they found it easier to lead employees into given courses of action. In general, if the judgments of the supervisors who received the training is considered a valid indication, the results of the training were highly positive.

Turning then to the results of the survey of the employees themselves, a large majority of the employees (70 per cent) said that their supervisors had a better understanding of how the men performed their jobs, 67 per cent reported that their supervisors had a better understanding of them as individuals, and 65 per cent reported that they now had a better idea of what their supervisors expected in terms of job performance.

The over-all results of this training program strongly suggest that supervisors can be trained to feed back appraisal results to employees, but it cannot be done effectively with just a lecture an hour or two long. It requires careful training and supervised practice.

Further evidence of the value of the training program described above has been reported by Meyer and Walker.[37] They found that of 49 employees reporting on the results of performance-appraisal feedback

---

[37] H. H. Meyer and W. B. Walker, "A Study of Factors Relating to the Effectiveness of a Performance Appraisal Program," *Personnel Psychology*, 14 (1961), 291–298.

sessions with their supervisors, 21 had taken constructive action as a result. There was a reasonably high correlation, $r = .63$, between the feeling by the employee that the supervisor had handled the feedback discussion well and action toward self-improvement taken by the employee after the session.

A somewhat similar approach to feedback sessions has been suggested by McGregor,[38] who feels that conventional appraisal places the supervisor in the untenable position of judging the subordinate and then acting on this judgment. McGregor suggests that the employee and the supervisor plan the former's development for a specified time period—say, six months. This planning is done initially by the employee, who then sits down with the supervisor; together they agree on the details of the developmental plan. At the end of the agreed-upon time period, the employee reports back to his supervisor and gives him a summary of progress made. This idea goes considerably beyond performance appraisal (or merit rating) as such, but it certainly looks as if it would be well worth serious consideration as an adjunct to the rating process.

An investigation that shows conventional appraisal to be considerably better than McGregor feels it to be has been reported by Mayfield. Mayfield,[39] who is Assistant Director of Industrial Relations at the Owen Illinois Glass Co., reported that 90 per cent of the employees interviewed at this company expressed satisfaction with the appraisal interviews. Only two per cent checked "no" and six per cent "undecided" in answer to the question: Should these discussions be continued?

One investigation that clearly showed how a failure to feed back rating results to the employees is related to job performance has been published by Huttner and O'Malley.[40] The study dealt with salesmen for a dental supply company who were working out of 21 branches. These 21 branches were chosen from 40 branches operated by the company; they were selected so that one group (10 branches) was the highest in sales and the other group (11 branches) was the lowest in sales. The manager of each branch rated each of his salesmen on a Performance Report that involved seven characteristics:

Over-All Job Performance
Motivation
Salesmanship
Organization and Planning
Job Knowledge
Advancement Potential
Attitude

---

[38] D. McGregor, "An Uneasy Look at Performance Appraisal," *Harvard Business Review*, 35 (1957), 89–94.

[39] H. Mayfield, "In Defense of Performance Appraisal," *Harvard Business Review*, 38 (No. 2, 1960), 81–87.

[40] L. Huttner and T. R. O'Malley, "Let Them Know!" *Personnel Psychology*, 15 (1962), 179–186.

Each salesman then rated himself on these factors. The effective and ineffective salesmen in each branch were identified from actual sales records. The study revealed that the ineffective salesmen rated themselves *almost exactly the same* as the effective salesmen in each of the seven characteristics.

The study further revealed that the low branch managers had rated their ineffective salesmen significantly lower in over-all job performance, motivation, and advancement potential. Since the effective and ineffective salesmen were not significantly different in aptitude test scores, age, training or experience, a conclusion of the investigation was that the ineffective salesmen were ineffective because they did not have accurate knowledge of their own weak characteristics. Huttner and O'Malley suggest that the supervisors of the ineffective salesmen should have given them a better idea of their weak as well as strong points and by doing so, i.e. by "letting them know," their job performance might quite possibly have improved.

Appraisal Interviews During regular work activities there are, of course, many opportunities for a supervisor to discuss with his subordinates certain aspects of their work, toward the end of helping them continually to improve their work performance, where this seems in order. A number of organizations, however, have established plans for supervisors to have appraisal interviews with their subordinates, in some cases on a regular basis, such as once or twice a year. Such interviews (also referred to as coaching) are difficult for some people to conduct, since they may (and frequently do) involve saying unpleasant things to the person being interviewed. (In fact, the interview might not serve its purposes unless, in one way or another, an individual achieves some insight into his deficiencies.) Where such interviews are to be carried out, supervisors preferably should be given training to help them to perform this function more effectively. (The study by Moon and Hariton[41] discussed earlier dealt with an evaluation of such a training program.)

It has been pointed out by Maier[42] that appraisal interviews may actually have various, and even conflicting, objectives, such as: (a) to let the subordinates know where they stand; (b) to recognize their good work; (c) to point out how and where they can improve; (d) to develop them in their present job; (e) to develop and train them for higher jobs; (f) to let them know how they may progress in the company; and (g) to warn some employees that they must do better. The way in which an interview is conducted should of course depend upon the purpose, and also upon the individual being interviewed.

It is not feasible here to discuss appraisal interviewing intensively,

---

[41] Moon and Hariton, *op. cit.*

[42] N. R. F. Maier, *The Appraisal Interview* (New York: John Wiley & Sons, Inc., 1958).

but the reader is referred to such sources as Maier.[43] He describes three types of appraisal interviews, namely the *Tell and Sell,* the *Tell and Listen,* and the *Problem-Solving.* The nature of these types is fairly well implied by their labels. He points out that each has its own uses—and its advantages and limitations. For example, the *Tell and Sell* method might well be appropriate with new and young employees who are inexperienced and insecure, and who want the assurance of an authority figure. The disadvantages of such an approach, however, are fairly obvious. It can result in stifling upward communication, and runs the risk of developing "yes" men. In the *Tell and Listen* method the supervisor typically starts off by covering the strengths and weaknesses of the subordinate. The second part is devoted more to thoroughly exploring the subordinate's feelings about the evaluation, the supervisor assuming a rather non-directive role during the interview. While such an interview may have positive value to the subordinate in getting things off his chest, it may not result in any positive, constructive program for the individual. The *Problem-Solving* interview is more directed toward establishing a mutual interest between the interviewer and the subordinate with the view toward employee development in a constructive, problem-solving manner with little emphasis on the appraisal as such. One problem with this approach is the possible conflicting dual role of the supervisor as both "judge" and helper.

With regard to interviewing techniques, Hoppock [44] gives 17 "ground rules" for conducting appraisal interviews. These "ground rules" cover such things as evaluating the man in terms of his actual job performance *on his job, listening* frequently to the man, considering the man's strengths as well as his weaknesses, not discussing salary or promotion during the interview, and letting the employee know in a tactful way exactly where he stands. The other "ground rules" are guides for seeing that the interview is conducted in a realistic and reasonable manner and that the integrity of the employee is respected.

In the discussion of appraisals as such, there is sometimes a question as to what information should be provided to the individual about his actual ratings, especially if they are expressed in numerical terms. Since ratings are not perfectly reliable, moderate (or even large) differences in ratings may not actually be reliable—as in comparing ratings from one time to another, or ratings of two people. Thus, it is usually desirable to avoid giving employees any numerical ratings. This difficulty can be largely eliminated by giving out general classifications such as A, B, C, and D. A "B-grade" employee is less likely to experience a change of grade as a result of a few numerical points difference in his rating than if he is classified exactly according to his numerical standing.

---

43 Maier, *op. cit.*

44 R. Hoppock, "Ground Rules for Appraisal Interviews," *Personnel,* 38 (No. 3, 1958), 31–34.

III

The Organizational and Social Context of Human Work

Human work is carried out in the context of some organizational and social environment. The individual becomes a part of this context, subject to whatever features it may have. Some of the aspects of this environment are planned and organized; some are the consequence of fortuitous events and circumstances. Included in this organizational and social context is whatever training is provided by the organization to persons who work within it. In addition, the characteristics of the organization impinge upon the individuals within it, as, for example, the structure of the organization, the type of supervision, the policies and the managerial practices that are used, the social environment (including

interaction with fellow workers), the communications (formal and informal), etc. In addition, the organization provides some set of incentives, including financial incentives. Thus the influence of the job-evaluation procedures on the wage and salary system is another characteristic of the organizational context of human work.

10

Training

The effective operation of any enterprise or system requires that the individuals involved learn to perform the functions of their current jobs at a satisfactory level of proficiency. An effective organization, however, also requires that it have available within itself a pool of individuals who are qualified to accept increasing responsibilities—to move into other jobs (usually at a higher level) with reasonable facility. This requires the development of personnel in areas that may not be important in their present positions but that may be important in future positions. Thus, an organization needs to provide opportunities for the continuous development of employees, not only in their present jobs, but also to develop their capabilities for other jobs for which they might later be considered.

The learning underlying such development can take place in either of two ways: through everyday work experience, or as the consequence of systematic training programs. In a broad sense, training can be considered to be any planned, organized effort that is specifically designed to help individuals develop increasing capabilities. Granting that the day-to-day "unplanned" learning through job experience is itself desirable (and certainly should not be discounted), nonetheless the experience of most organizations has supported the conviction that it is advantageous

to plan systematic training programs of various types as a regular part of an adequate personnel development program. Such programs are definite assets in such ways as helping employees to learn the correct job methods, to achieve a satisfactory level of job performance, and to acquire capabilities that would be valuable in possible future jobs. For certain types of material to be learned, organized training may result not only in more *efficient* learning, but also in *more* learning than the unplanned learning through normal experience.

To achieve its purposes, however, a training program should be established on sound principles and practices that are conducive to human learning. While many training programs have been established on such principles and practices, some have not been. The ineptness of some programs probably can be attributed in part to the bandwagon aspects of training in the last two or three decades. Because of the upsurge of training programs during these years, some companies, operating on a keep-up-with-the-Joneses philosophy, have established training programs simply because it was "the thing to do," and not on the basis of fulfilling a real need within their organizations.

THE PURPOSES OF TRAINING

The nature and purposes of training programs vary a great deal, but the purposes generally can be grouped into three categories:

Developing Job Knowledge and Skills Many programs are directed toward developing knowledge and skills that will be useful to employees in performing either their present jobs or possible future jobs. Usually such training is job-oriented. A particular comment might be made regarding "skills." While this term frequently is viewed as covering manual skills, it is suggested that it be considered in a much broader context, including such skills as those involved in interpersonal relations, supervision, organization and planning activities, and other related abilities.

Transmitting Information The purpose of some training programs is directed toward transmitting general, rather than job-oriented, information, such as information about the company, its products and services, and its organization and policies.

Modifying Attitudes The intent of some training programs is that of changing the attitudes of employees in various ways, such as developing more favorable attitudes among employees, increasing motivation, and increasing the sensitivity of (especially) supervisory and management personnel to the feelings and reactions of other people.

CHARACTERISTICS OF LEARNING

Since the crux of any training program is predicated on how adequately people learn, it would be useful to us, before discussing training as such, to consider the characteristics of learning and some of the factors that are associated with the learning process. The topic of learning has been studied extensively by psychologists, especially by experimental psychologists and educational psychologists. While most studies of learning have been carried on outside the industrial context, it is probable that many of the results of such studies, and especially the methods of investigation, have reasonable applicability to the industrial situation.

The Nature of Learning Learning has been defined by McGehee and Thayer [1] as a description of behavioral changes which result from experience. It is impossible to observe directly the process that we call learning. The fact that learning has occurred can only be inferred from a comparison of an individual's behavior prior to and subsequent to experiences of specific kinds. This is not to say that there has been no "learning" if there is not an overt behavioral change, but this cannot be *known* unless there *is* a behavioral change. Thus, operationally, we must define learning as involving an observable behavioral change. While the nature of the changes within the organism are as yet unknown, the behavioral changes that occur typically can be observed, and it is these changes that can be systematically investigated.

A point might be added regarding these behavioral changes. In characterizing human behavior, it is frequently the practice to talk in terms of a stimulus (S) acting upon an organism (O) to bring about a response (R). The purpose of training is to "establish the connection" between given stimuli and desirable responses. In a job situation, the "inputs" to any individual are the stimuli that he senses: communications, instrument readings, materials, equipment, etc. Given any input (or combination), there is presumably some output that would best achieve the objectives of the job; the responses an individual makes are his "output." Thus, training is the means by which individuals learn to respond appropriately to their input stimuli.

Learning Curves The behavioral changes that occur as the consequence of learning frequently can be quantified and presented in the form of a learning curve. Such a curve typically shows the cumulative changes that occur during the learning period, either on the part of individuals or (when averaged for several individuals) on the part of groups. A learning curve shows the level of achievement (in terms of an

---

[1] W. McGehee and P. W. Thayer, *Training in Business and Industry* (New York: John Wiley & Sons, Inc., 1961), p. 132

appropriate criterion) over the learning period. The criteria will of course depend upon the circumstance, but may be measures of productivity, performance on tests, or otherwise.

The form and length of learning curves varies considerably from one situation to another. The total period of training on some jobs, for example, may be months or even years, whereas on other jobs it may be only days or weeks. Further, the shape of the curve may vary. Fig. 10.1

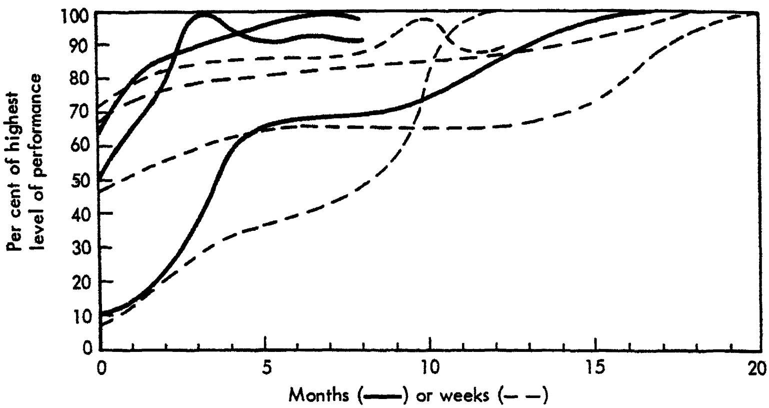


Fig. 10.1. Illustrations of generalized learning curves for several different jobs. While the original curves were somewhat irregular, for illustrative purposes they have been smoothed.

illustrates several generalized forms of learning curves. While these represent generally the curves of several actual jobs, the natures of the jobs are not particularly pertinent to the central point of illustrating the differences. These curves are averages for groups of employees, and the criterion scales for all of them have been converted to a common, arbitrary base. Another learning curve was shown in Chapter 2 (Fig. 2.9).

The curve for any single individual can be similarly drawn. In fact, a very practical use for learning curves is that of comparing the curve for an individual with that for a group. For such purposes, however, it is preferable to include a curve that also includes the ranges of performance for some segment of the total illustrative group, such as the 25*th* and 75*th* percentiles, or the group that is represented by a standard deviation above and below the mean (these including about 67 per cent of the cases). Such curves are illustrated in Fig. 2.9 (Chapter 2). The curve for an individual, when compared to such a figure, makes it possible to determine how well the individual is progressing during his training,

relative to others. If an individual falls markedly below the average, it may be in order to consider giving him special attention, or possibly to consider his transfer to another type of work.

With reference to learning (and learning curves) it is generally the case that the *relative* degree of improvement in learning is greater in the case of more difficult jobs than in the case of easier jobs. This is illustrated by the comparisons shown in Fig. 10.2. Each part of this figure shows the

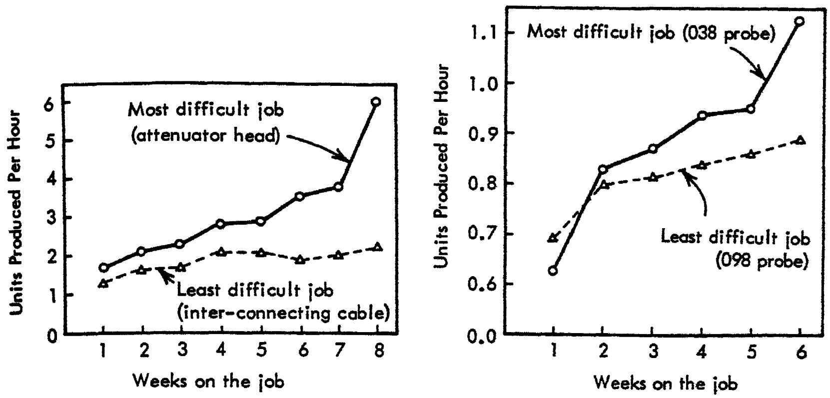


Fig. 10.2. Comparison of learning curves of two pairs of related jobs that differ in difficulty level. Note that the *relative* improvement is greater in the more difficult jobs. (Courtesy of Tetronix, Inc., and Dr. Guyot Frazier.)

production (in units per hour) throughout training of employees on two pairs of related jobs in the manufacture of oscilloscope accessories; in each pair, one job had been judged by management representatives as "most difficult" and the other as "least difficult." The greater relative improvement in the case of the most difficult job is evident in each pair. This difference can generally be attributed to the fact that an "easy" job is one for which most people already have the basic acquired skills and knowledges in ready-to-use form with little adaptability of these required. In the case of the more difficult jobs, this is not the case. The development and integration of the required skills and knowledge must be carried out over a period of time.

The Plateau In learning any complex task it often happens that, after a certain level of efficiency has been attained, a period of time arrives in which little or no improvement takes place. This period is followed by a later increase in skill. The period during which no apparent improvement occurs is known as a *plateau*. Plateaus are not always present in the learning curves for industrial operations, and when they are present it is sometimes possible to eliminate them by changed or improved

methods of instruction. No matter what method of instruction may be in use, however, arrival at a plateau in the learning process often is characteristic of the average learner's performance.

Examples of plateaus in learning are shown in a couple of curves of Fig. 10.1. While the explanation for a plateau may not be apparent, it is important for the trainer to be aware of it in order that he can give assurance to the trainees (when they reach a plateau) that this is "par for the course," and that they can expect later additional progress. Such information can aid in preventing the trainees from becoming discouraged with the apparent leveling in their progress. By analyzing the activities and training procedures it may be possible for the trainer to diagnose the reasons for a plateau, and possibly to modify the training to minimize it. Whether he is able to do this or not, his awareness of the fact that this situation is by no means uncommon in learning processes will aid him to cope with the situation in a more intelligent manner.

Retention of Learning In the case of learned material, it is usually the case that people tend to lose (to forget) some of it over time if the information (whatever it may be) is not used, or if it is not supplemented with additional learning. The extent and rate of forgetting varies with the type of information, the extent to which it was originally "learned" (and the extent of "over-learning," if any), and the individual. Generally the relearning of material takes less time than the original learning.

In connection with the learning of perceptual-motor skills, however, there is some fairly current evidence to suggest that retention of proficiency over time may be fairly high. This inference comes from a study by Fleishman and Parker.[2] In this experiment the subjects were trained on a tracking device, and were then brought back in groups of about 10 for retraining at varying time intervals after original training, the time intervals being 1, 5, 9, 14, and 24 months. The results are summarized in Fig. 10.3; this shows the learning during the original learning period and also the average level of performance during the later relearning sessions. The two curves represent different learning groups—those who were given formal guidance versus those who learned "on their own." In the case of both groups, the performance during the relearning period was virtually the same as at the end of the original learning period. While not shown on the figure, it should be added that there were virtually no differences in performance between those whose time span since original learning was short (1, 5, or 9 months) versus long (14 or 24 months).

---

[2] E. A. Fleishman and J. F. Parker, Jr., "Factors in the Retention of and Relearning of Perceptual-Motor Skill," *Journal of Experimental Psychology*, 64 (1962), 215–226.

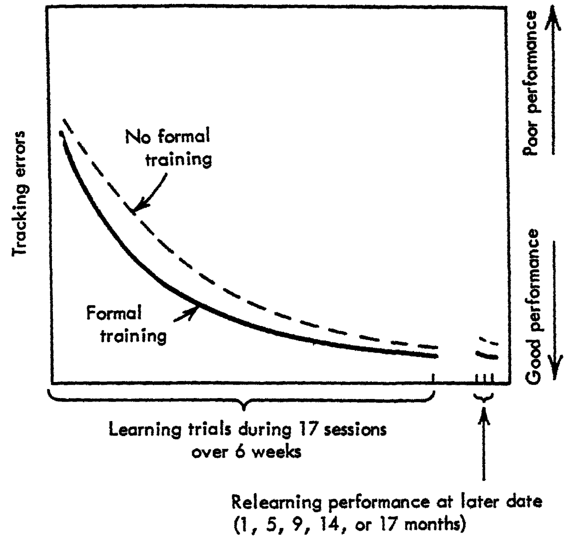


Fig. 10.3. Comparison of performance on a perceptual-motor task during original learning period and during a relearning period at a later time. Note that the performance during the relearning period was about the same as at the end of the original learning period. (Adapted from Fleishman and Parker, *op. cit.*)

THEORIES OF LEARNING

Over a period of several decades, various theories of learning have been proposed by different psychologists. A theory has been described as a "way of binding together a multitude of facts so that one may comprehend them all at once." [3] Thus, a learning theory would be one that is intended to explain in conceptual terms what takes place when learning occurs, how the learning takes place, and what variables facilitate the learning process. To be valid, such a theory must then account for all of the types of human learning behavior that occur, whether in raising children, in school, in learning to play golf, or in learning a job in industry. As of this date, however, no single theory has evolved that has been generally accepted as being completely adequate in these terms.

It is not within the province of this book to describe and evaluate the various learning theories.[4] It may be useful, however, to mention some of the generalizations of the theories postulated by the theorists that

---

[3] G. A. Kelley, *The Psychology of Personal Constructs*, Vol. I (New York: W. W. Norton & Company, Inc., 1955).

[4] For a discussion of learning theories the reader is referred to such sources as E. R. Hilgard, *Theories of Learning*, 2nd ed. (New York: Appleton-Century-Crofts, 1956).

have implications for industrial training as presented by McGehee.[5] On the basis of a synthesis of various theories he points out that certain generalizations seem to be common to different theories (though presumably to varying degrees). These generalizations may be considered as statements of "what happens" when an individual learns. Thus, they can serve as guideposts for ordering experiences that are planned to modify behavior, i.e., train the individual. These generalizations are given below:

1. The learner has a goal or goals, i.e., he wants something.
2. The learner makes a response, i.e., he does something to attain what he wants.
3. The response, which he makes initially and continues to make in trying to attain what he wants, is limited by: (a) the sum total of his past responses and his abilities; (b) his interpretation of the goal situation; (c) the feedback from his responses, i.e., the consequences of his response.
4. The learner, having achieved his goal (or goal substitute), can make responses which prior to his goal seeking he could not make. He has learned.

The implications of these generalizations are that the learning situation must be structured within the framework implied by them. Thus, the goals of training must be those which are meaningful to the learner (although these are amenable to modification), the learner must be provided with the opportunity to make appropriate responses (to "do" something, to take an active rather than a passive part), the responses required must be within the constraints specified in (3) above, and there must be assurance that the learner has learned the desired behaviors (which implies continual evaluation of the training).

While, in a sense, the learning theorists and the practical trainers are poles apart, McGehee argues for more and more interchange between them, toward the end of bringing to bear upon the practical problems of training the benefits that can accrue from the application of greater understanding of the process of human learning.

PRINCIPLES OF LEARNING

While there is much that is not known about human learning, and while current theories are not in complete agreement with each other, nevertheless, there have evolved certain operational principles of learning that presumably have some applicability to training processes. It must be pointed out, however, that certain questions have been raised by

[5] W. McGehee, "Are We Using What We Know About Training?—Learning Theory and Training," *Personnel Psychology*, 11 (1958), 1–12.

Gagné [6] regarding the generality of some of these. Further mention will be made of his views later. A few of the most widely accepted principles will be discussed.[7] It will be noted that some of these relate back to the generalizations from theories mentioned earlier.[8]

Knowledge of Results One such principle deals with knowledge of results. There actually are two ways in which this principle can have an effect upon learning. In the first place, it provides the basis for correcting one's errors. There are some kinds of tasks in which such information is virtually mandatory for learning. A crane operator, for example, would have trouble learning to manipulate the controls of a crane appropriately unless he knew how the crane responded to his control actions. The second way in which knowledge of results can have a facilitating effect is by making the task more interesting to the learner; it can thus have a motivational effect.

The effects upon learning of knowledge of results have been confirmed in a number of studies. One such study dealt with rifle practice.[9] Without describing the details, the situation was one in which men fired rifles at moving targets in a simulated rifle range on each of ten days. Where "knowledge of results" was to be furnished to the subjects, there was an electronically activated "check" that informed the subject that he had made a hit. The subjects were divided into two equivalent groups, one group receiving knowledge of results, the other not, for the 400 "runs." After that, the knowledge of results was removed from the group that previously had received such feedback. The results, shown in Fig. 10.4, indicate a distinct difference between these groups during the first 400 trials. After that, however, (and aside from a brief spurt) the first group levelled off in its performance while the second group (with no knowledge at any time) continued to improve, and finally caught up with the first group. It took nearly twice as long to reach the same level of performance.

Another study of the effects upon learning of knowledge of results is one reported by Lindahl [10] in the training of cutters of tungsten discs. In this operation, a pedal controls the grinding wheel which cuts discs off the end of a tungsten rod. A paper tape recording of the foot action was used as a means of feeding back knowledge of the results of the pedal action; see Fig 10.7 later in this chapter. This method of feedback

---

[6] R. M. Gagné, "Military Training and Principles of Learning," *American Psychologist*, 17 (No. 2, February, 1962), 83–91.

[7] For a review and summary of factors associated with learning see R. M. Gagné and R. C. Bolles, "A Review of Factors in Learning Efficiency," *Automatic Teaching*, E. Galanter, ed. (New York: John Wiley & Sons, Inc., 1959).

[8] McGehee, *op. cit.*

[9] H. C. W. Stockbridge and B. Chambers, "Aiming, Transfer of Training, and Knowledge of Results," *Journal of Applied Psychology*, 42 (1958), 148–153.

[10] L. G. Lindahl, "Movement Analysis as an Industrial Training Method," *Journal of Applied Psychology*, 29 (1945), 420–436.

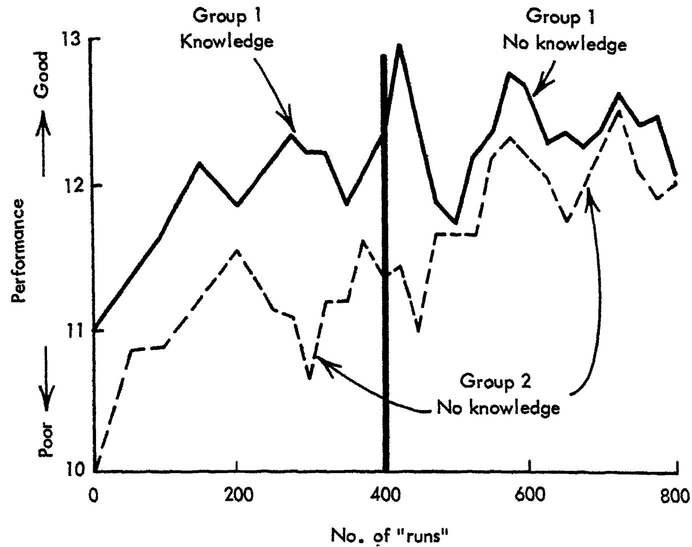


Fig. 10.4. Illustration of effects of knowledge of results in a learning task. This particular task was rifle practice. Note the leveling off of learning on the part of group 1 after knowledge of results was removed. (Adapted from Stockbridge and Chambers.)

had a significant effect in speeding up the learning process and in reducing the number of broken grinding wheels.

In connection with the feedback of knowledge to trainees, there are certain considerations that should be kept in mind in providing for the feedback. In the first place, the knowledge about his performance that is fed back to the trainee should be specific rather than general.[11] The more specific the knowledge (within reason), the more rapid the improvement and the higher the level of performance. In the second place, one should not overburden the trainee with information. Only that information should be presented that is appropriate to the level of proficiency of the trainee at the time, and that information should be at the level of specificity that is optimum at that time. And in the third place, the feedback about a trainee's performance should take place as soon as possible after the performance. Its effect will be dissipated in relation to the time lag.

The evidence regarding the effectiveness of knowledge of results in learning is so general that there should be provision in virtually every

[11] R. B. Ammons, *Knowledge of Performance; Survey of Literature, Some Possible Applications, and Suggested Experimentation*, USAF, Wright Air Development Center, Technical Report 54–14, 1954.

training program for giving feedback to trainees about their performance.

Motivation One of the generalizations postulated by McGehee [12] on the basis of most learning theories is that the learner has a goal or goals, i.e., he wants something. While there are some theoretical questions about the importance of motivation in learning, there is general agreement that a motivated learner learns better than an unmotivated one. Chapter 12 will discuss motivation further, but it can be pointed out here that motives generally can be characterized as *internal* and *external*.[13] Intrinsic motivation is related to the task itself; there is some direct relationship between the task and the goal of the learner, such as in the case of a mechanic who achieves satisfaction from a job well done. Extrinsic motivation is independent of the task, i.e., the task is viewed as a possible means to some other end, such as for the income the job might ultimately bring to the learner. Intrinsic motivation, of course, has an advantage over extrinsic motivation, and can provide the trainee with a continuing job interest even after completion of training. But in the case of new employees on many jobs in industry, there may be no initial basis for intrinsic motivation on the part of the trainees. Where this is the case, it is up to the trainer to do what he can to generate the basis for the ultimate development of intrinsic motivation, and also to provide incentives that can stimulate external motivation. Some such incentives have been suggested by McGehee and Thayer.[14] These include: praise (sincerely put and from a valued person); good working conditions; pleasant relations with peers, supervisors, and subordinates; and status within the work group and community. Other incentives include earnings (or the prospects thereof), the recognition of the need for training, respect for the trainer, and job security.

Reward versus Punishment With reference to motivation, a question might be raised as to the relative effectiveness of positive rewards (for desirable behavior) versus negative punishment (for undesirable behavior) in learning. Without bringing in supporting evidence, it has been concluded that, in general, punishment is less effective in learning than is reward.[15] It has also been pointed out that punishment (especially if severe) can serve to fix the undesirable behavior rather than to eliminate it, and that it may have some undesirable by-product effects such as dislike of the punishing person.[16] On the other hand, it has been pointed out that mild punishment may be quite effective, if administered immediately following the incorrect response, and if it is informative.[17]

---

12 McGehee, *op. cit.*

13 E. R. Hilgard, *Introduction to Psychology*, 3rd ed. (New York: Harcourt, Brace & World, Inc., 1962), p. 326.

14 McGehee and Thayer, *op. cit*, p. 145.

15 Hilgard, *op. cit.*, p. 328.

16 *Ibid.*

17 McGehee and Thayer, *op. cit.*, p. 160.

Such information should be specific to the incorrect response only, and clearly differentiated from concurrent responses that are correct.

Operant Conditioning As a special aspect of the use of incentives in learning, particular mention should be made of the concept of "operant conditioning" as developed by Skinner.[18] This refers generally to the strengthening of a stimulus-response relationship by following the (desired) response with a "reinforcing stimulus," which is some sort of "reward." Without going into the various ramifications of this, it can be said that the crux of Skinner's formulation is that of providing fairly immediate reinforcement (reward) for desirable behavior, and not reinforcing other behavior. The "reinforcement schedule," i.e., how frequently the response is reinforced, has some effect upon the acquisition and retention of the desired response.

In the practical training situation, the operant conditioning principles would argue for frequent reinforcement (reward of some type) for the behaviors that it is desired that the trainees learn. The nature of the "reward," however, then becomes the question. A very practical reinforcement is the use of frequent praise and approval for satisfactory performance. In some types of situations it may be possible to work into the learning task some fairly automatic procedure for assuring the trainee of the adequacy of his performance. This is the essential basis for programmed instruction procedures that will be discussed later in this chapter.

Distribution of Learning Periods In many learning situations the length and spacing of learning periods are important for efficient learning. Generally speaking, people learn more effectively when the learning is split up into periods than when it is scheduled continuously. The optimum lengths of periods and of interspersed changes (either rest or other activity) vary with the material to be learned. A study on the distribution of learning on an industrial job is reported by Mahler and Monroe.[19] The training in question was job instruction training to supervisors, designed to qualify the supervisors to be more effective in giving job training to new employees. One group of 300 supervisors received six hours of training spread over a two-week period. Another group received the six hours in three two-hour sessions on successive days. In a later follow-up it was found that the first group (for which the training was spaced over two weeks) made fewer mistakes in training new workers than did those in the second group (for which training was concentrated in three days).

---

[18] B. F. Skinner, "The Science of Learning and the Art of Teaching," *Harvard Educational Review*, 24 (1954), 86–97.

[19] W. R. Mahler and W. H. Monroe, *How Industry Determines the Need for and Effectiveness of Training*, Personnel Research Section, Department of the Army, PRS Report No. 929, March 15, 1952.

The principle has been fairly well established that learning efficiency is enhanced when the learning periods are of optimum duration and are spaced appropriately. The problem for the trainer, however, is figuring out what the optimum duration and spacing of learning periods should be for a given type of training. While there are at present no solid guidelines to propose, two or three general comments might be made. In the first place, it would seem desirable to have each training period cover some cohesive segment of training content. Further, one should avoid the onset of excessive boredom or inattention by interspersing breaks now and then. (Most trainers can get some cues about lag of attention by observing their captive audience.) In addition, it has been found that even short breaks frequently facilitate the learning process.

Whole versus Part Learning Still another type of variable that has been found to influence learning is that of "whole" versus "part" learning. This refers to the size of the "chunks" of training content that are presented to the trainees. There is a fair amount of support for the proposition that learning tends to be facilitated when the content of a given presentation forms a reasonable and meaningful whole, rather than an assortment of little bits and pieces than are covered individually during training. But one can also cite evidence to support the contention that "part" training is superior. The implication of such conflicting evidence is that "it all depends"—that is, it depends upon the particular situation (the job, the types of trainees, etc.). In the development of training programs one presumably should not then blindly follow the practice of "part" versus "whole" training, but rather should use that approach that seems most appropriate to the specific situation. While there are no easy, cookbook rules for the trainer to follow in this regard, some factors to be considered have been proposed by McGeoch and Irion [20] and subsequently paraphrased by Hilgard.[21] All of these factors have been found to be important in one or more situations:

1. The more intelligent the subject, the more likely that the whole method will prove advantageous.
2. The advantage of the whole method increases with practice in using it.
3. When practice is distributed rather than massed, the whole method becomes increasingly favorable.
4. Material that is meaningful and unified tends to favor the whole method.
5. The total length of the material, the actual sizes of the parts, and the number of parts making up the whole must be considered. (There is no simple rule here, for it is quite possible that very short and very long passages will profit from the whole method, with in-between ones favoring the part method.)
6. A disadvantage of the part method is the time required to connect the

---

[20] J. A. McGeoch and A. L. Irion, *The Psychology of Human Learning,* 2nd ed. (New York: Longmans, Green & Co., Inc., 1953), pp. 501–507.

[21] Hilgard, *op. cit.,* p. 314.

separately learned parts. Methods that get around this difficulty will reduce the advantages of the whole method.

7. Following the separate learning of the parts, attempted recall may reveal more mutual interference among the parts with some materials than with others. Hence the disadvantage of the part method may depend upon the material to be learned.

Transfer of Learning Transfer of learning refers to the extent to which that which is learned in one situation can be carried over to, or be applicable to, another situation. Training frequently is carried out in something of an artificial situation, rather than in the real job situation. This is the case with job training that is not "on the job" as such, many kinds of simulators (such as of aircraft), management games, conferences, and role-playing techniques, as well as other forms of training. Where any off-the-job training takes place, the assumption is made (at least implicitly) that the learning that takes place in the training situation will transfer to the real job situation. Aside from the training situation, the question of possible transfer of learning also comes up when an individual changes from one job to another, especially if there is some kind of similarity between the two jobs.

Psychologists have given a great deal of attention to the study of transfer of learning, beginning with the very early work of Thorndike and Woodworth.[22] The conclusion reached by these early investigators—and substantiated by several more recent studies—is that the amount of transfer from one skill or mental function to another is far less than was formerly believed. Indeed, Thorndike [23] has concluded that transfer in a general way does not occur at all and that what is often regarded as transfer is simply due to *identical elements* in the two jobs under consideration. These identical elements may be activities (usually manipulative activities in industry), or methods of work. An example of the former would be any routine assembly operation that is identical with the operation in some other assembly job. An example of the latter would be the methods of laying out work that would be similar from one job to another in the activities of a motion study or layout man.

As applied to training situations, this theory would suggest that the transfer of learning from the training situation to the job would depend upon the extent to which there are identical elements common to the two. In a study by Frisby,[24] for example, it was found that an aircraft training device (specifically, a landing machine trainer) facilitated learn-

---

[22] E. L. Thorndike and R. S. Woodworth, "The Influence of Improvement in One Mental Function upon Efficiency of Other Functions," *Psychological Review*, 8 (1901), 247–261, 384–385, 553–564.

[23] E. L. Thorndike, "Mental Discipline in High School Studies," *Journal of Educational Psychology*, 15 (1924), 1–22, 83–98.

[24] C. B. Frisby, "Field Research in Flying Training," *Occupational Psychology*, 21 (1947), 24–33.

ing in landing and take-off, but did not in any way carry over to other aspects of flying. In other words, on the basis of the identical-elements theory, one presumably cannot expect more transfer of learning to occur than can be expected in terms of definite similarities between the training situation and the actual job (or between one job and another).

In connection with the transfer of learning, however, there are some nagging questions that are not adequately accounted for by the postulated theory of identical elements, and that as yet remain unanswered. In the first place, questions have been raised about the general applicability of the transfer-of-training concept as based on identical elements. Further, if one does accept this concept, how does one determine which elements are identical? And if the two elements are somewhat similar but not identical, how much transfer could one expect? These and related questions are very real to those responsible for training programs, and for the development of training equipment. In connection with the increasing use of simulators and many kinds of training devices, for example, there is considerable current interest in the question of the "fidelity" of the devices. This, in effect, refers to the degree of similarity of the training devices to the equipment and facilities used on the job. Here one needs to distinguish between the degree of *physical* fidelity and that of *psychological* fidelity. Psychological fidelity refers to the degree of similarity of the human operations and activities that are involved. It is the psychological fidelity that is critical in the transfer-of-learning context. While the psychological fidelity typically would be optimum when there is also a high degree of physical fidelity of the training situation, a high degree of physical fidelity is by no means a universal requirement for psychological fidelity to occur (although it may frequently be desirable that some critical features of a training device be similar to those of the real equipment). To cite a hypothetical (and somewhat obvious) example, if one were using a mock-up of the control room of a refinery as a training device, it probably would not matter a whit that the mock-up was of wood and the "real thing" of metal; but the operational aspects of the mock-up presumably should be very comparable to those of a real control room itself.

Since the original concept of identical elements has not been found to be the "key" to account for transfer-of-learning effects (at least not in a practical sense), others have formulated suggested theories. Among these are the formulations of Gagné, Baker, and Foster,[25] and that of Osgood.[26] These (and other) theories tend to be somewhat more analytical,

---

[25] R. M. Gagné, Katherine E. Baker, and Harriett Foster, *On the Relationship between Similarity and Transfer of Training in the Learning of Discriminative Motor Tasks,* USN, Special Devices Center, Technical Report SDC 316-1-5, 1949.

[26] C. E. Osgood, "The Similarity Paradox in Human Learning; a Resolution," *Psychological Review,* 56 (1949), 132-143.

and, in particular, are concerned with the "degree" of similarity of the stimuli and of the responses between the training and the actual situations. Osgood[27] has developed a "transfer surface" to characterize the relationships that he has postulated. This is shown in Fig. 10.5. For any

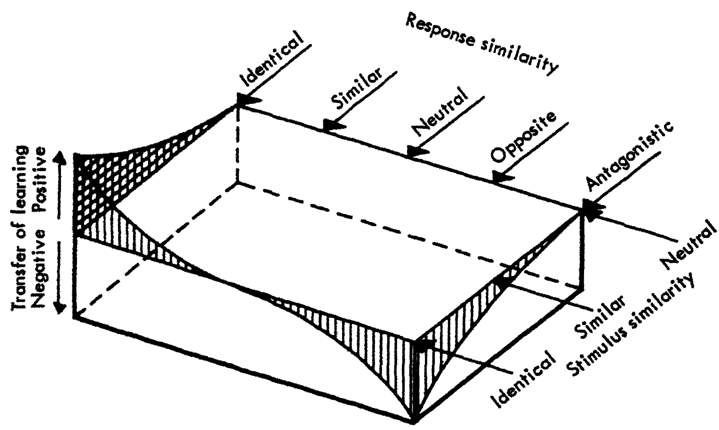


Fig. 10.5. The "transfer surface" suggested by Osgood, *op. cit.*, to demonstrate the relationship between degree of stimulus similarity, response similarity, and transfer of learning (amount and direction).

given combination of stimulus similarity (between the training and real situations) and of response similarity, the amount (and direction) of transfer is indicated by the vertical scale. Transfer would be greatest where the stimulus and the response during training are identical with those of the job. Where both are unrelated (neutral), no transfer would be expected. Where the stimulus is identical but the response is antagonistic, the transfer would be expected to be negative. This frame of reference, then, leads one to think in terms not of identical elements but rather in terms of the degree of similarity, and also leads one to look at the training situation in terms of both stimuli and responses that are involved—and their degree of relationship with their counterparts in the job itself.

It should be pointed out that questions have been raised regarding certain aspects of Osgood's formulation. For example, Gagné and Bolles[28] question the implied requirement of response similarity and express the opinion that response similarity is a factor concerning which we know very little. It should also be noted that Osgood's model is not quantitative in nature; how does one "measure" the degree of similarity, say, of

[27] Osgood, *op. cit.*

[28] Gagné and Bolles, *op. cit.*

stimuli? It is also suggested that the formulation may be more applicable to training situations where specific stimuli and corresponding specific responses can be identified, as opposed to training in concept formation, problem solving, and other higher-order mental operations. But granting its possible limitations, it does seem to crystallize a tentative frame of reference for considering transfer of learning in training situations.

Generalizations about Human Learning Aside from the above principles of learning, there are some other generalizations regarding learning processes that can be made. Hilgard,[29] for example, has set forth 14 statements about learning that seem to be close to the level of common sense, but that have considerable factual evidence to support them While there is no unanimity of opinion about these, Hilgard suggests that there would be substantial agreement about the applicability of these in learning situations:

1. In deciding who should learn what, the capacities of the learner are very important. Brighter people can learn things less bright ones cannot learn; in general, older children can learn more readily than younger ones; the decline of ability with age, in the adult years, depends upon what it is that is being learned.
2. A motivated learner acquires what he learns more readily than one who is not motivated. The relevant motives include both general and specific ones, for example, desire to learn, need for achievement (general), desire for a certain reward or to avoid a threatened punishment (specific).
3. Motivation that is too intense (especially pain, fear, anxiety) may be accompanied by distracting emotional states, so that excessive motivation may be less effective than moderate motivation for learning some kinds of tasks, especially those involving difficult discriminations.
4. Learning under the control of reward is usually preferable to learning under the control of punishment. Correspondingly, learning motivated by success is preferable to learning motivated by failure. Even though the theoretical issue is still unresolved, the practical outcome must take into account the social by-products, which tend to be more favorable under reward than under punishment.
5. Learning under intrinsic motivation is preferable to learning under extrinsic motivation.
6. Tolerance for failure is best taught through providing a backlog of success that compensates for experienced failure.
7. Individuals need practice in setting realistic goals for themselves, goals neither so low as to elicit little effort nor so high as to foreordain to failure. Realistic goal-setting leads to more satisfactory improvement than unrealistic goal-setting.
8. The personal history of the individual, for example, his reaction to authority, may hamper or enhance his ability to learn from a given teacher.
9. Active participation by a learner is preferable to passive reception when learning, for example, from a lecture or a motion picture.

---

[29] Hilgard, *op. cit.*, 1956.

10. Meaningful materials and meaningful tasks are learned more readily than nonsense materials and more readily than tasks not understood by the learner.
11. There is no substitute for repetitive practice in the overlearning of skills (for instance, the performance of a concert pianist), or in the memorization of unrelated facts that have to be automatized.
12. Information about the nature of a good performance, knowledge of his own mistakes, and knowledge of successful results, aid learning.
13. Transfer to new tasks will be better if, in learning, the learner can discover relationships for himself, and if he has experience during learning of applying the principles within a variety of tasks.
14. Spaced or distributed recalls are advantageous in fixing material that is to be long retained.

Discussion As indicated earlier, however, some serious questions have been raised by Gagné[30] regarding the relative utility of some of the commonly accepted principles of learning as applied to the practical problems of job training. He cites chapter and verse to illustrate that such "principles" as distribution of practice, reinforcement, and response familiarity have not been found to be effective in certain job-training situations. He indicates, rather, that the more effective principles (in at least the cases cited) were those of a very different nature, as follows:[31]

1. Any human task may be analyzed into a set of component tasks which are quite distinct from each other in terms of the experimental operations needed to produce them.
2. These task components are mediators of the final task performance. That is, their presence insures positive transfer to a final performance, and their absence reduces such transfer to near zero.
3. The basic principles of training design consist of: (a) identifying the component tasks of a final performance; (b) insuring that each of these component tasks is fully achieved; and (c) arranging the total learning situation in a sequence which will insure optimal mediational effects from one component to another.

Gagné points out, however, that such "principles" as task analysis, intra-task transfer, component task achievement, and sequencing do not deny the *relevance* of the traditional principles of learning, but rather raise questions about their *relative importance*. While these newly proposed principles are not yet supported by any well organized body of experimental evidence, it would seem that the implied focus of attention on the job tasks and operations that are to be performed would contribute to the pinpointing of training toward specific, meaningful job objectives.

Tasks vary in their natures, however. Gagné and Bolles,[32] for ex-

---

[30] Gagné, *op. cit.*
[31] Gagné, *op. cit.*
[32] Gagné and Bolles, *op. cit.*

ample, characterize many job tasks as falling into one of the following general categories:

Identification (of objects, materials, controls, tools, etc.)
Following procedures
Using concept (applying principles)
Motor skills

Given a task that is to be learned, the approach to the training of personnel must be adapted to the nature of the task. Thus, a procedural task involves the learning of sequence of task components; a trouble-shooting task, the learning of certain rules and the identification of, say, test instruments; an assembly operation, the learning of a sequence of job elements and the development of appropriate manipulative skills; an accounting task, the learning of certain rules and procedures, etc. It is postulated here that, for job tasks of any given "type" (such as learning a sequence of procedural steps), some particular basic "method" of training probably could be identified that would contribute most effectively to learning of tasks of that type. Some such methods very likely would draw upon certain of the traditional learning principles and generalizations. At the present time, however, there is no organized basis for specifying *what* principles or practices would be best for any given type of task. This is one of the current important challenges to the learning psychologists who are interested in industrial training problems.

TYPES OF TRAINING PROGRAMS

To give an overview of the gamut of training programs in industry, some of the more common types of programs will be described briefly.

Orientation Training These training programs are directed toward orienting new employees (and in some cases present employees) to the company. In particular, it is for the purpose of providing such information about the company as its organization, its history, its products or services, its policies, its procedures, and its personnel. Such information may be conducive to the development of favorable attitudes on the part of employees, and may also help new employees to become more effective in their jobs because of their knowledge of such matters as policies, procedures, and organizational relationships.

On-the-job Training A great deal of training is carried out on the job itself. Where new employees are being trained, they may be given instruction in some cases by their supervisors, in other cases by an instructor. The amount of attention given to training may be very nominal and informal, or it may be organized and formal. On-the-job training has such advantages as training on the actual job situation, which virtually eliminates the possible problem of transfer of learning.

On the other hand, however, where such training is treated very casually and is viewed as a nuisance by the supervisor, its effects may leave something to be desired. Such training can be very effective, however, if it is appropriate in terms of the nature of the job, if it is given in the form of a systematic, well formulated program, and, preferably, if it is given by a separate job trainer rather than by the supervisor.

Off-the-job Training Sometimes job training cannot well be given on the job, as, for example, if new untrained employees might damage equipment or materials, injure themselves or others, or if the nature of the training itself would unduly hamper on-going operations. The off-the-job training may be carried out in a simulated job situation with equipment and facilities that are similar to or identical with that used on the job. Vestibule training schools are examples. In such training, the emphasis is rather strictly on training, and not on production as such. Most types of supervisory and management training programs are off-the-job training. Such programs may be directed toward improving the effectiveness of such personnel in their present jobs, or toward developing them for possible future jobs.

Frequently off-the-job training is combined with on-the-job training to provide an appropriate balance, and to facilitate transfer of the training to the actual job situation. Apprentice programs typically are of this type.

Other Training Programs Among other training programs is the type that is used by some companies in training college graduates who have been employed upon graduation; such training frequently consists of a period of months, or even years, of experience in many of the departments of the company. Although such employees usually are assigned to jobs in whatever department they are with at the time, this variety of experience usually is thought of as training for future management responsibilities, by giving them an intimate familiarity with many phases of the company. Another type of training carried out by some companies provides opportunities for employees to take general subjects (that may, or may not, be related to their jobs), and cultural, and even recreational subjects.

DETERMINING TRAINING NEEDS

In connection with the practical aspects of industrial training, there are three basic considerations to which the trainer needs to give his attention. These are:

What are the training needs?
What method of training should be used?
How should the training be evaluated?

Let us now turn our attention to the first of these: What are the training needs?

As implied before, training should be directed toward the two-fold objectives of developing personnel to perform their present jobs (or new jobs) effectively, and of developing personnel so that they can progress to other jobs. The disparity between the *current* level of proficiency of personnel on the one hand, and the *desired* level on the other hand, characterizes the *training needs.* The primary mission of a training department is that of narrowing that gap. With continual personnel changes and organizational changes, this virtually becomes a never-ending process. The first step in filling that gap is, of course, that of determining what the training needs actually are.

For purposes of our discussion we might separate this problem into two parts (although in reality these two blend into each other). In the first place, there is the need to train employees on new jobs, especially new employees. In the second place, there is the need to identify deficiencies on the part of current personnel that could be compensated for through training; this is essentially an on-going employee development effort.

Training Needs: Job Training Where inexperienced personnel are to be trained for new jobs, the training "needs" are fairly manifest, namely, those of helping the individuals to acquire the knowledge, skills, and attitudes that are required in the performance of the jobs in question. This requires a detailed analysis of the job; further mention will be made of this later.

Training Needs: Employee Development In the case of continuing efforts toward general employee development, the identification of training needs has a somewhat different focus: that of identifying the specific deficiencies of employees in performing their on-going activities. In the case of employee development for possible future jobs, the requirements of the future jobs must be identified. It should be pointed out here that this process can be concerned with *individual* employees or with *groups* of employees. In the case of an individual, the intent is to determine in what respects he needs additional training; this is what McGehee and Thayer[38] refer to as *man analysis.* In this connection, if there is an employee evaluation program, the evaluation of individuals may be very useful. Various methods of employee evaluation were discussed in Chapter 9 and need not be repeated here.

If, on the other hand, there are indications that the employees of some group (for example, those on some given job) require additional training, the intent would be that of identifying the specific deficiencies that characterize the group. Thus, a program might be developed for,

---

[38] McGehee and Thayer, *op. cit.*

say, salesmen or supervisors. For such purposes, the methods of determining training needs range from those which are entirely subjective to those which are quite objective. The nature of some training situations is such that the determination of the needs must of necessity be based on subjective judgments. A survey by Mahler and Monroe[34] dealt in part with the analysis of training needs of 150 companies. In the portion of the survey relating to methods of determining training needs, the training directors of the companies were asked to indicate their preferences for various methods for production, clerical, and supervisory and technical employees. A summary of these expressed preferences is given in Table 10.1. Although there are variations for the three types of employees, it can be seen that, generally speaking, the most common methods preferred were (a) "informal observation," (b) requests from management, (c) talks with supervisors, (d) group discussions, and (e) analysis of reports.

It is evident from the preferences for methods listed in Table 10.1 that the determination of training needs at present is predominantly based on subjective judgments or observations. Although we sometimes

Table 10.1

PER CENT PREFERENCES OF 150 TRAINING DIRECTORS FOR VARIOUS METHODS OF DETERMINING TRAINING NEEDS

Method of determining training needs	Type of employees: Productive	Clerical	Technical-supervisory
a) Informal observation	12	12	11
b) Requests from management	17	16	15
c) Talks with supervisors	16	17	14
d) Talks with non-supervisors	4	6	3
e) Group discussions	10	10	13
f) Training advisory committees	5	5	6
g) Questionnaires to supervisors	2	2	2
h) Questionnaires to trainees	1	2	4
i) Supervisory morale survey	1	1	4
j) Employee morale survey	4	4	2
k) Tests	3	4	2
l) Merit ratings	4	6	4
m) Intensive interview with supervisors	2	4	4
n) Interview with union officials	2	1	4
o) Analysis of reports (costs, turnover, grievances, etc.)	12	7	7
p) Other	4	3	3
Total	100%	100%	100%

Source: *Mahler and Monroe*, op. cit.

[34] W. R. Mahler and W. H. Monroe, *op. cit.*

are inclined to discount the validity of human judgments or observations (frequently with good cause), it should be kept in mind that sometimes the *method* of obtaining judgments or observations affects their validity. In our present context of determining training needs, for example, there have been at least a few training circumstances in which some interesting techniques have been developed.

Check list of training needs. Two illustrative activity check lists for recording "observations" have been presented by Fryer, Feinberg, and Zalkind.[35] These check lists are shown in Tables 10.2 and 10.3. Table 10.2 is part of a check list to be used by a training specialist in observing the work of a production department. It provides for indicating by a "yes" or "no" check whether or not each check list statement has been observed in the department. As filled out for a hypothetical situation, it can be seen that certain observations suggest possible training needs. The other check list (Table 10.3) is one that would be applicable for identifying, in a similar manner, the possible training needs for employees on a particular type of job, in this specific illustration the job of supervisor.

Judgments of training needs. A systematic approach to obtaining judgments of employees regarding training needs of management personnel was developed by Wood.[36] The study will be described in some detail—more to illustrate the method than the results as such.

Table 10.2

CHECK LIST FOR NEEDED TRAINING

Items recorded by training specialist	*Is item seen in this production unit?* YES	NO	*Possible training need*
Downward communications to supervisor slow	x		x
Frequent gripes by supervisors about tools		x	
Supervisors use suggestions from workers		x	x
Many rejects returned to production		x	
Customers' complaints from sales ignored		x	
Turnover in shop higher than seems necessary	x		x
Many lost-time accidents		x	
Paper work up to date		x	x
Raw-material delays		x	
Supervisors use specialists of company		x	x

Source: *Fryer, Feinberg, and Zalkind,* op. cit.

[35] D. H. Fryer, M. R. Feinberg, and S. S. Zalkind, *Developing People in Industry* (New York: Harper & Row, Publishers, 1956).

[36] W. F. Wood, "Identification of Management Training Needs" (Ph.D. thesis, Purdue University, June, 1956).

To begin with, about 500 management personnel were asked to list or describe their own training needs, or those of their bosses. These listings were used to develop three questionnaires, each with 68 items. The questions in each questionnaire were similar, except for the frame of reference of the three questionnaires, as follows:

1. Your personal need.
2. Your superior's need.
3. Your subordinate's need.

Table 10.3

SAMPLE CHECK LIST FOR SUPERVISORS

Items recorded by training specialist	*Checked for adequate performance* YES	NO	*Possible training need*
Keeps inventory of tools	x		
Prepares training outline for apprentices		x	x
Takes unsafe machinery out of service	x		
Checks all repairs	x		
Maintains "hours of work" record	x		
Inspects regularly for quality of product		x	x
Informs on elimination of waste		x	x
Plans workplace layout		x	x
Instructs on cost of materials		x	x
Explains company policy to workers		x	x

Source: *Fryer, Feinberg, and Zalkind*, op. cit.

The items in the questionnaires dealt with seven major areas of training needs as follows (one item from each area is given to illustrate the types of items on the questionnaire):

Area	*Example of item*
1. Subordinate relations.	Coaching my subordinates to help them improve their performance.
2. Personal development.	Getting my ideas down on paper.
3. Labor relations.	Handling a grievance.
4. Wage and salary.	Understanding and applying the company's wage payment policy.
5. Management control.	Understanding the application of departmental budgets.
6. Technical knowledge.	Further technical information on manufacturing processes.
7. Knowledge of company.	How the company products are sold to the consumer.

Each item could be checked by indicating whether an individual felt that he (or his immediate supervisor or subordinate) had "No

need," "Some need," or "Real need" for training as indicated by the item.

The statistical analysis of a preliminary try-out with the questionnaires indicated fairly satisfactory reliabilities of the responses for the seven areas (with six of the seven reliability coefficients being .86 or above). The final forms of the questionnaires were administered to management personnel in three plants of a large company, as follows:

Form	*No.*	*Position*
Your personal need	310	Gen foremen and superintendents
Your superior's need	824	Foremen and gen. foremen
Your subordinate's need	89	Superintendents and top mgmt.
Total	1,223	

One type of analysis made with the resulting data dealt with the degree to which training needs were perceived by *individuals* (self perceptions) as contrasted with their needs as perceived by their *superiors* and by their *subordinates.* Fig. 10.6 illustrates the pattern that emerged

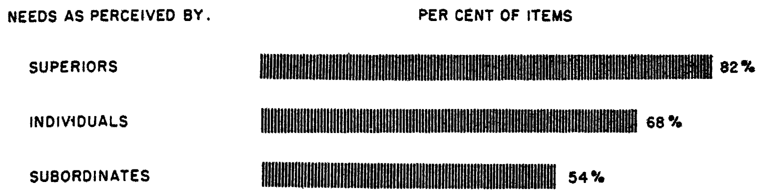


Fig. 10.6. Per cent of training needs items in "management control" area for which need was indicated by individuals, by their superiors, and by their subordinates. (From Wood, *op. cit.*)

in all seven training-needs areas, this particular example being that of management control. In this, and in all the other areas, the *superiors* perceived the training needs of individuals whom they supervised as being greater than did the individuals themselves. In turn, the *subordinates* of the individuals perceived the training needs of those individuals (their superiors) as being *less* than did the individuals themselves.

Discussion The identification of training needs is a critical phase of any training program. Too frequently in industrial organizations there is a vague impression that "we need more emphasis on training," or a tendency to hop on the training bandwagon simply because other companies are doing so. Such points of view are indeed not conducive to the careful analysis of the actual needs that a well-formulated training program might reasonably be expected to fulfill. An adequate training

program should proceed after valid, pinpointed training needs have been identified by the use of careful, analytical methods.

JOB ANALYSIS FOR TRAINING PURPOSES

Where the purpose of a training program is that of job training, some type of detailed study of the job itself is virtually mandatory, leading to the identification of the skills, knowledges, and attitudes that are required for successful job performance. Most typically, this involves two processes that have been characterized by Miller[37] as *task description* and *task analysis*.

Describing Tasks The description of a task involves a statement (or statements) of what Miller refers to as the *requirements* of the task in question. In other words the description describes the (usually) overt, observable activities involved in the task—the things a person "does." Usually, such activities are of a physical nature, although they may be of a communicative nature. Since the nature of tasks varies greatly, there is no universal pattern that they should follow, and no universal set of variables that should be specified. Where a task involves some fairly standard cycle of activities, however, it has been suggested by Miller[38] that task activity statements specify the *indication or cue* which which calls for a response, the *object* to be used (such as a control device that is to be activated), the *activation or manipulation* to be made, and the *indication of response adequacy* or feedback. Thus, such a task description would specify *when* the individual is to do something, *what* he is to do it with, what *action* he is to take, and what *feedback* there is to indicate that the action has been performed adequately. In effect, then, the task description characterizes the input and output aspects of the task. As a simple illustration let us take the task of opening an automatic door at a garage. An analysis of this task in the above terms might be something like that below.

Task: opening automatic garage door
- Indication or cue (when): sound of automobile horn
- Object (what): push button control
- Activation or manipulation (action): press push button
- Indication of response adequacy (feedback): observe door raised to overhead position, and hear motor stop.

While the analysis of many tasks would be much more complex than this, this example will at least illustrate the process.

---

[37] R. B. Miller, "Task Description and Analysis," in *Psychological Principles in System Development*, R. M. Gagné, ed. (New York: Holt, Rinehart, & Winston, Inc., 1962).

[38] Miller, *op. cit.*

Specifying Training Requirements With the parameters of the required inputs and outputs, it is then in order to identify the knowledges, skills, and attitudes that are required to bring about the "transformation" between the inputs and outputs. As pointed out by Crawford,[39] these are psychological in nature, such as sensing, discriminating, remembering, deciding, and choosing. The identification and specification of these are what Miller [40] refers to as *task analysis.*

The bridge from task descriptions on the one hand, to the training requirements on the other hand (the required skills, knowledges, and attitudes), however, is not a simple process, as is effectively pointed out by Altman.[41] While the "breakthrough" in such methodology that Altman looks forward to has not yet materialized, the bridge is one that needs to be established in actual training situations. Until the "breakthrough" occurs, the transition must be made by the careful analysis and good judgment of those who are able to infer from the task descriptions what skills, knowledges, and attitudes are required to perform the tasks, and to express these in behavioral terms that have psychological meaning.

Special Types of Job Analysis While task description and analysis is the most common method of deriving training requirements for jobs, in special situations it is possible to develop techniques that have unique applicability to the job in question. An example of such a unique technique is illustrated by the work of Lindahl.[42] Although the particular method he used was specific to the job in question, this example illustrates how a bit of ingenuity can be useful in developing a method appropriate to a particular circumstance. He set up a simple graphic recording apparatus to show the stroke pattern followed in cutting tungsten discs. The machine studied made use of a rotating abrasive wheel that is lowered by means of a foot pedal to cut off a disc. After the disc has been cut, the foot pressure is released, the abrasive pressure is relieved, the abrasive wheel rises, the tungsten rod advances, and the machine is ready for a repetition of the cutting cycle. The recording apparatus used by Lindahl is illustrated in Fig 10.7. Identification of the various parts of the cutting cycle is shown in Fig. 10.8.

The cutting patterns found at various stages of training in the performance of a typical trainee are illustrated in Fig. 10.9. This partic-

---

39 M. P. Crawford, "Concepts of Training" in *Psychological Principles in System Development,* R. M. Gagné, ed. (New York: Holt, Rinehart, & Winston, Inc., 1962).

40 Miller, *op. cit.*

41 J. A. Altman, "Methods for Establishing Training Requirements," in *Uses of Task Analysis in Deriving Training and Training Equipment Requirements,* USAF, WADD Technical Report 60–593, December, 1960.

42 L. G. Lindahl, "Movement Analysis as an Industrial Training Method," *Journal of Applied Psychology,* 29 (1945), 420–436.

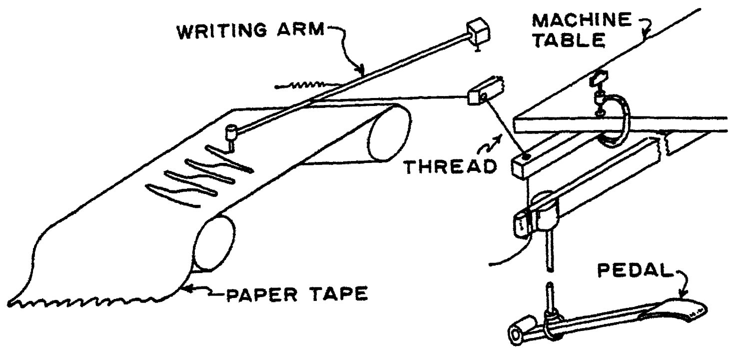


Fig. 10.7. Schematic drawing of apparatus used in obtaining graphic records of job performance in cutting discs with a foot-operated abrasive wheel.

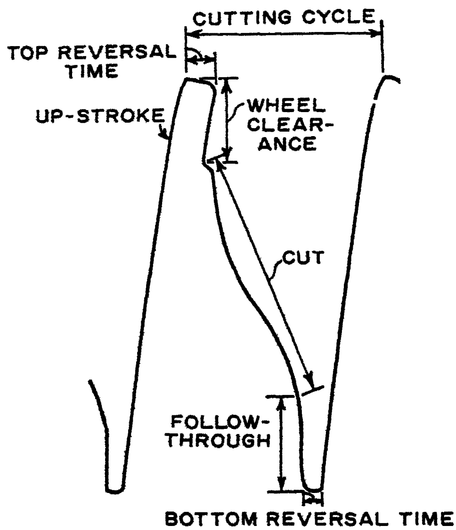


Fig. 10.8. Identification of various parts of the cutting cycle as shown by the graphic record attained with the apparatus shown in Fig. 10.7.

ular method of analyzing the job also lent itself to use as a training device. By using this simple equipment to show trainees the specific errors they were making in their cutting cycles, Lindahl was able to improve the quantity and quality of work at various stages of training, and also to increase the life of abrasive wheels.

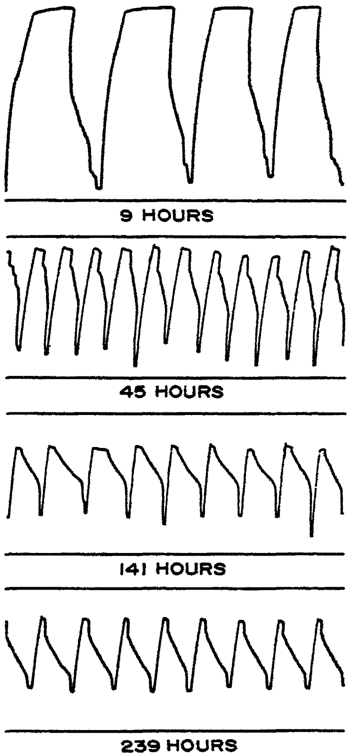


Fig. 10.9. Records of performance of a trainee at various stages of his experience on the job.

TRAINING METHODS AND AIDS

Many different types of training methods and techniques are used, including adaptations for specific purposes. Some of the methods and techniques are listed below:

1. Lecture
2. Demonstration
3. Films and TV
4. Programmed instruction
5. Simulators and training aids
6. Conference
7. Sensitivity training
8. Case study
9. Role playing

10. Management games
11. Other methods

Selection of Training Method In the development of training programs, some determination has to be made as to what specific *method* or *technique* should be used for a particular *training objective*. Any given method would be effective in certain training contexts, but not in others. As pointed out by McGehee and Thayer,[43] however, there is limited concrete evidence now available to guide those responsible for training in the selection of the most appropriate techniques for a specific training problem. They make a plea for carefully controlled research—first, to determine the relative efficiencies of various techniques with different types of training problems, and second, to find out what learning principles are involved in each. Until more comprehensive answers are available, experience and logical analysis, plus some inklings from research, must be used in the selection of training methods for specific situations.

While it is not feasible here to attempt any exhaustive analysis of the various methods or techniques, some of these will at least be described, and in certain cases some discussion will be added.

Lecture The lecture has been severely criticized as a method of training, primarily on the grounds that it does not normally provide for active participation on the part of the trainees; this lack of participation, in turn, precludes any feedback to them. While such criticisms would argue against its indiscriminate use, there are, however, various circumstances in which it is an appropriate method of training. It has been proposed for use, for example, under the following circumstances:[44] when presenting completely new material to a group; when working with a large group; when introducing another instruction method; when classroom time is limited; and when summarizing material developed by another instruction method.

Demonstration The demonstration technique is especially applicable in job training where some physical process is to be learned, particularly one that involves the use of equipment of some type. If the process itself is to be learned, however, it is probable that the demonstration itself should be followed by active participation and practice by the trainees.

Films and TV The use of films in instruction received quite a boost during World War II both in industry and the military services because of the heavy training demands at the time. TV is, of course, a more recent development. Obviously neither technique, as such, will compensate for poorly presented training material. Three points might be made about these techniques as they relate to the learning process. In the first

---

[43] McGehee and Thayer, *op. cit.*, pp. 193–194.

[44] J. H. Proctor and W. M. Thornton, *Training: A Handbook for Line Managers* (New York: American Management Association, 1961).

place, there are certain kinds of presentations that can be made more effectively by such methods than by any other. Cameras can sometimes be placed where a learner cannot be—as, for example, in demonstrating surgical techniques or certain mechanical operations. In addition, these techniques facilitate demonstrations (such as by animation) that otherwise might not be feasible. In the second place, there is no substantial evidence to suggest that such methods are *generally* more effective in presenting regular course material than other methods, but in some *individual* circumstances they have been found to be more effective. This was found to be the case, for example, in a study reported by Fryer [45] on three methods of training Air Force personnel in code learning. Four groups of men, equivalent in initial ability, were covered by the study. The training for the four groups was as follows:

1. Lecture (oral instruction with slides).
2. Manual (students read illustrated manual).
3. Film (15-minute animated film).
4. No training.

The content of the three training methods was the same.

Immediately after training, and again after two months, an objective 25-item test was given to the groups with results as shown in Table 10.4. The results in Table 10.4 show that the film method was superior, both immediately following training and after two months, to either of the other two methods investigated.

Table 10.4

TEST SCORES FOR FOUR AIR FORCE GROUPS TRAINED IN CODE LEARNING BY DIFFERENT METHODS

Method of Training	Average Test Scores Following Training	After Two Months
Film	17.9	16.3
Manual	15.4	13.0
Lecture	15.2	13.0
No training	5.4	7.0

Source: *Fryer*, op. cit.

Films have also been used successfully in certain types of industrial training. VanderMeer [46] reports a study in which sound films were used in the training of engine lathe operators. He found a significant superi-

[45] D. H. Fryer, "Training with Special Reference to its Evaluation," *Personnel Psychology*, 4 (1951), 19–37.

[46] A. W. VanderMeer, "The Economy of Time in Industrial Training: An Experimental Study of the Use of Sound Films in the Training of Engine Lathe Operators," *Journal of Educational Psychology*, 36 (1945), 65–90.

ority in the operators trained by this method when compared with those trained by traditional methods.

Even when films or TV are not superior to other methods, however, they may be more economical than having a lecturer, since the films or TV tapes can be used over and over again.

In the third place, a possible disadvantage of these methods is the characteristic lack of opportunity for participation by the learners. Participation sometimes can be provided for afterward, however, in such ways as by having discussion groups.

Programmed Instruction In recent years there has been a great deal of interest in programmed instruction both in education and in industry. That which we will call programmed instruction also goes by several other labels: teaching machines, automated training, auto-instruction, and programmed learning. Because of the increasing importance of programmed instruction in industry, several aspects of it will be discussed.[47]

Programmed instruction has been described by Cook and Mechner [48] as the application of the science of learning to the tasks of training and education, and as being characterized by the following elements:

1. Active response by the student.
2. Small steps in which careful control of stimuli produces gradual increments in mastery of the subject.
3. Immediate feedback for each response.
4. Self-pacing, or individualization, of the rate at which the learner masters the material.
5. Low error-rate for the individual learner, as a consequence of the effective operation of the first four principles.

Perhaps the central features of programmed instruction as far as psychological principles of learning are concerned are those of participation (active response) by the learner, and immediate feedback to him.

Background The earliest significant effort in this area was that of Pressey,[49] who developed a simple mechanical device for use in testing students. It contained a window through which a multiple-choice test item was presented, and a keyboard with four keys corresponding with the four

---

[47] The reader interested in further material relating to programmed instruction is referred to such sources as:

S. Margulies and L. D. Eigen, eds., *Applied Programmed Instruction* (New York: John Wiley & Sons, Inc., 1961).

A. A. Lumsdaine and R. Glaser, *Teaching Machines and Programmed Learning; A Source Book* (Washington, D.C.: National Education Association of the United States, 1960).

F. F. Kopstein and I. J. Shillestad, *A Survey of Auto-Instructional Devices*, USAS, ASD Technical Report 61-414, Sept. 1961 (available for purchase from Office of Technical Services, Dept. of Commerce, Washington, D.C.).

[48] D. Cook and F. Mechner, "Fundamentals of Programmed Instruction," in Margulies and Eigen, *op. cit.*

[49] S. L. Pressey, "A Simple Apparatus Which Tests and Scores—and Teaches," *School and Society*, 23 (No. 586, March 20, 1926), 373–376.

possible responses. Only when the key of the correct answer was pressed did the next question appear. Thus, feedback was made available immediately.

While Pressey really laid the groundwork for programmed instruction, it was the later research of Skinner [50] relating to operant conditioning that probably sparked the recent flurry of interest in this technique.

Variations in Programmed Instruction Methods Most of the essential features of programmed instruction can be embodied in any of a number of different methods of presentation. These include elaborate teaching machines (some of which are tied in with computers) that present material visually, on films, or on sound tapes; more simple machines (but that still involve electronic or mechanical features); simple devices that hold the paper on which the "program" is printed (some with a manual procedure for uncovering the "frames" successively); programmed books; and (in very simple form) illustrations, diagrams, and other printed material.[51] Examples of two devices are shown in Fig. 10.10.

Whatever method of presentation is used—elaborate or simple as the case may be—it performs at least three functions: (1) it presents information and/or questions or problems to the learner; (2) it provides for him to respond; and (3) it provides feedback as to whether the response is right or wrong.

The mechanics of performing these three functions are in one sense incidental; they are means to an end. The critical feature of programmed instruction is the programming itself. It should be pointed out, however, that computer-based instructional systems, and certain other types of teaching machines, make possible the programming of material that otherwise might not be possible.

Programming The typical instructional program used in programmed instruction consists of a series of "frames" (also referred to as "images," "items," or "pages"). Each frame deals with a small segment of information that is intended to provide for some sort of response on the part of the learner. The nature of the frames varies from one program to another, but examples are given in Figs. 10.11 and 10.12.[52] The frames are ordered in sequence, and typically one frame builds upon the previous frame. As the learner progresses from frame to frame (receiving feedback in each case), he then builds up, bit by bit, the total subject matter of the program.

---

[50] Skinner, *op. cit.*

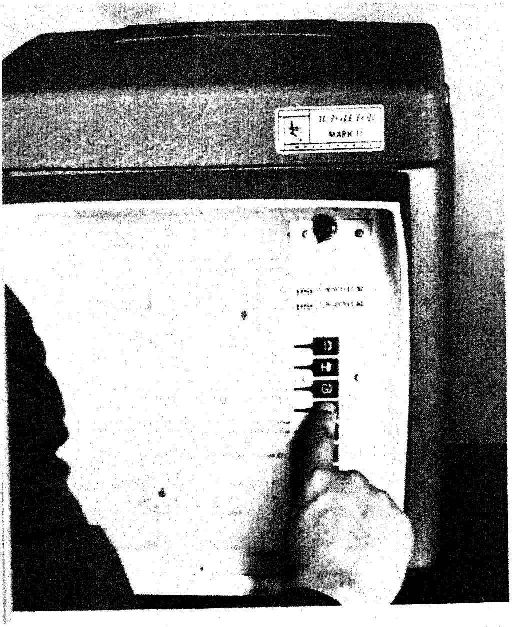
[51] Various types of devices are discussed or illustrated in the following:

Margulies and Eigen, *op. cit.*

Kopstein and Skillestad, *op. cit.*

W. L. Ross, Jr., *et al., Teaching Machines: Industry Survey and Buyers Guide* (New York: The Center for Programmed Instruction, Inc., 1962).

[52] J. P. Lysaught and C. M. Williams, *A Guide to Programmed Instruction* (New York: John Wiley & Sons, Inc., 1963).



A.

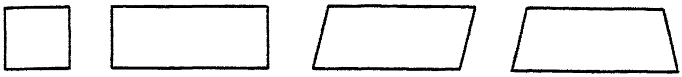
B.

Fig. 10.10. Examples of two programmed instruction devices.

A. The Autotutor Mark II (Reg. U.S. Pat. Off.); courtesy Educational Science Division of U.S. Industries (12345 New Columbia Pike, Silver Spring, Maryland).

B. The Didak 506; courtesy Rheem Califone Corp., P.O. Box 78567, Los Angeles, California.

Q. Quad means four. Lateral refers to side. A quadrilateral always has four sides. A square would be one type of quadrilateral. The figures below are all ____________ because they have four sides.



R. quadrilaterals

Q. A rectangle  is a quadrilateral because it always has _________ sides.

R. four

Q. A rectangle is one type of _____________________ because it has _________ sides.

R. quadrilateral
four

Q. A square is a type of _________________________ because it has __________ sides.

R. quadrilateral
four

Q. All figures that have four sides are known as ________________

R. quadrilaterals

Fig. 10.11. Illustrative frames from a programmed instruction sequence. This particular sequence involves very small "steps." The correct response (shown as R) is not in view of the respondent until he has entered his own response. (From Lysaught and Williams, *ibid.*)

There are certain differences in the approach to programming that should be mentioned. In the first place, the responses can be "constructed" or "multiple choice" in nature. A constructed response is one in which the learner actually writes in his response; it requires recall, or the formation of an answer based on what he has learned. A multiple-choice response is one in which the learner selects one of several possible responses, and requires recognition rather than recall. While opinions differ regard-

S. The hotter the filament, the brighter the lamp. The cooler the filament, the less ____________ the lamp.

R. bright

S. When the voltage is held constant, an increase in resistance results in the decrease in current; naturally, then, a decrease in resistance leads to an ____________ in current.

R. increase

S. The + sign stands for addition; the ÷ sign stands for d________n.

R. division

S. A proper noun names a particular person, place or thing. John, Austria, Rochester, Genesee River are all proper nouns. New York is a ____________ noun because it is the name of a ____________ state.

R. proper
particular

S. River is not a ____________ noun because it is not the name of a particular person, place or thing. Hudson River is a ____________ noun.

R. proper
proper

Fig. 10.12. Illustrative programmed instruction frames from various programs. Only one or two frames are shown from each program; thus, these examples do not illustrate the sequence of frames within a single program. The correct response (shown as R) is not in view of the respondent until he has entered his response. (From Lysaught and Williams, *ibid.*)

ing the desirability of one over the other, the more active participation required with constructed responses would seem to argue in its favor, whereas the multiple-choice form facilitates the "mechanics" of the instruction, such as the feedback to the learner.

Another distinction that is of some consequence is that between

"linear" versus "branching" programs. In the case of linear programs, each subject goes through all frames in sequence, and he must master each one in turn. It is the form that is proposed by Skinner.[53] A "branching" program is one that makes it more possible to adapt to the level of achievement of the learner. This is done by providing, at specific frames in the program, "branches" to be followed, such as by those who have not adequately mastered the material to the point in question. This is the practice followed by Crowder.[54] The branching technique is used in various "scrambled texts," which are books that have been developed on programmed instruction principles. If, on the basis of a learner's response, he is to take one of the "branches" of the program, he will be instructed to refer to the appropriate page.

In connection with some of the differences in programming practices, it is probably too early to indicate what practices are more conducive to effective learning or to learning certain kinds of material. Considerable additional research will be required to provide these answers.

It should be stated that programming is a time-consuming process, and requires individuals who are well qualified to carry it out. In developing an industrial program, for example, it was estimated that it took between 30 and 60 minutes per frame.[55] When one considers the fact that a program can take hundreds or thousands of frames, it can be recognized that programs are not simply pulled out of a hat; they require time and effort.

Uses in industrial training. While programmed instruction was initially used in educational situations, it was not long before its possible applications to industrial training became apparent. While the use of this procedure in industry is fairly limited as yet, there are a number of companies that have adopted the procedure, and are using it. It has been estimated that, in general, such programs reduce training time by about two-thirds.[56]

As an example of the use of programmed instruction in industry, Hickey [57] reports the case of a training program in a mail-order business for the job of package billing; this is the final clerical operation in the preparation of packages to be sent to customers. The original training program for this job covered a period of 40 hours. A program was developed for programmed instructional procedures, with the material being

---

[53] B. F. Skinner, "Teaching Machines," *Science*, 128 (1958), 969–977.

[54] N. A. Crowder, "Automatic Tutoring by Means of Intrinsic Programming," in Galanter, *op. cit.*

[55] J. P. Lysaught, "Programmed Learning and Teaching Machines in Industrial Training," in Margulies and Eigen, *op. cit.*

[56] W. L. Ross, Jr., "The Industrial Market for Programmed Instruction," in Margulies and Eigen, *op. cit.*

[57] A. E. Hickey, "Programmed Instruction in Business and Industry," in Margulies and Eigen, *op. cit.*

incorporated in workbooks. In the study, 120 new trainees were used, with 4 groups of 15 being given the regular training and 4 groups being given the programmed-instruction workbooks. The average time taken by the programmed-instruction groups was 26 hours, resulting in a saving of 34 per cent over the 40 hours of the regular instruction procedures, with no impairment in job performance. Considering the fact that the company hired 500 girls for this operation every year, there would be a saving of $20,000 of salaries and overhead if they were trained in this manner.

As another example, programmed instruction was used in the training of telephone relay adjustors.[58] Since the job depends largely on manual skills, the program required that the instructions be read to the trainees by the use of magnetic tape recordings in combination with a filmed visual presentation. A comparison of the cost of training of relay adjustors trained by the original method and by programmed instruction is shown in Fig. 10.13. The area under each curve represents the value of the production of the trainee during the first year. The area above each curve represents the cost of training in terms of lost production. The dif-

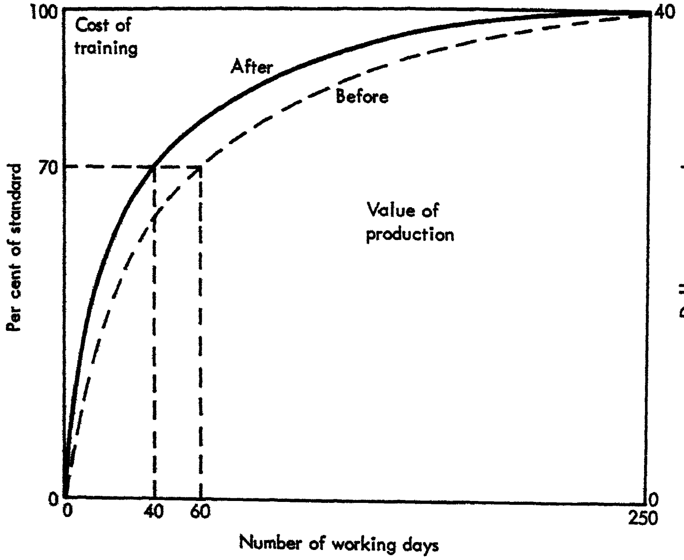


Fig. 10.13. Cost of training an apprentice relay adjuster before and after programmed instruction. The difference between the two curves represents the savings in cost of training relative to the value of production. (From Hickey, *ibid.*, p. 291.)

[58] Hickey, *ibid.*

ference between the two curves represents the savings by the programmed-instruction method. With programmed instruction, the time to reach 70 per cent of standard was 40 days, as contrasted with 60 days by the original method. Taking into account all costs, including supervision, it was estimated that the total savings for a group of 40 adjusters would be about $41,000, a reduction of about a third of the previous training cost of $120,000.

Discussion Programmed instruction has generally been viewed as an aid to instructors, to be used in combination with other training methods and techniques. There are those, however, who forecast the development of teaching machines to the point that they would virtually replace other forms of instruction.[59] In contrast to such views, Pressey[60] views with considerable alarm the "boom" trend of teaching machines. Aside from certain criticisms regarding some of the principles and practices of programmed instruction, he makes a strong case for viewing such instruction as "adjunct autoinstruction," which would aid in the use (and increase the value) of other instructional methods and materials.

It is not up to us to forecast the long-range developments (although we are inclined toward Pressey's point of view). It is certainly safe to say, however, that there is no possibility within the reasonable future of training personnel becoming automated out of their jobs. It also seems reasonable to suggest that there is a tremendous potential for the increased use of such procedures in many phases of industrial training, perhaps as adjuncts to other training methods and techniques. While one can generally view the increased use of programmed instruction with enthusiasm because of the possible training benefits, a couple of restraints are in order. In the first place, an organization should not adopt such procedures simply because other companies are doing so. Rather, programmed instruction should be adopted as a training method only if and when there is reason to believe that it is the best method for the purpose, taking into account the improvement in learning and the matter of time and costs. And in the second place, the time and cost considerations are formidable factors. The cost of program development (which can run into thousands of dollars) and the cost of whatever equipment is involved make it feasible only if the costs can be spread over a substantial number of personnel. In this connection, at least one industry is developing programs that can be used on an industry-wide basis, thus spreading the development costs over the participating companies.

But, while complete programs may not be economically feasible in specific situations, it is still possible on a limited basis to apply some of

---

59 Galanter, *op. cit.*, Chap. 1.

60 S. L. Pressey, "Teaching Machine (and Learning Theory) Crisis," *Journal of Applied Psychology*, 47 (1963), 1–6.

the principles of programmed instruction to those segments of training programs for which they seem especially appropriate, such as in the preparation of manuals and related materials.

Simulators and Training Aids Simulators and training devices are used to provide trainees with physical equipment that resembles to some degree the equipment that is to be used on the job. Usually such devices are used when it is impractical for some reason to use the actual equipment (such as possible injury to the trainees or others), or when the cost of the actual equipment is excessive. Such devices range from simple mock-ups, models, and prototypes, to extremely complex simulators such as those used by the military services for, say, training aircraft pilots.

As indicated earlier in the discussion of transfer of training, however, the potential utility of such devices depends more on the psychological fidelity than on the degree of physical fidelity. In terms of their utility in training, it is critical that they reproduce with reasonable accuracy the aspects of the real job that are central to the process of transfer of training. In other words, the various stimuli and responses, and the intervening mental operations and decisions that are involved in the training situation, should correspond reasonably well with those of the job itself.

Aside from simulators and training devices as such, there are many other types of training aids that can be developed or obtained. Film and TV, for example, can be viewed as training aids. In addition, one can use charts, blackboards, flip-sheets, manuals, hand-outs, and scale models, to mention a few. Fryer, Feinberg, and Zalkind [61] suggest the following ten-item check list for use in evaluating training aids. This check list will help the trainer to select the most effective device:

1. *The aid must help the trainee to be more critical.* In the use of films and film strips, their dramatic appeal must not lull the trainee into passive acceptance of the subject-matter. The aid and the manner of presentation should stimulate independent and critical thinking.
2. *The aid must give an accurate picture of the concept which it represents.* A training aid must not distort the presentation. The scale must be accurate so as not to confuse or exaggerate the object.
3. *The aid must be suitable to the level of abilities and experience of the trainee.* A training aid must be selected that is neither too advanced nor too simple for the intelligence and educational background of the trainee.
4. *The aid and all its significant parts must be seen clearly by both small and large groups.*
5. *The aid must be easy to maintain and repair.* An aid that breaks down frequently and requires frequent readjustments will be too discouraging to use.

---

[61] D. H. Fryer, M. R. Feinberg, and S. S. Zalkind, *Developing People in Industry* (New York: Harper & Row, Publishers, 1956).

6. *The aid must be economical to ship or store.* Nothing on it should be so small or unique as to encourage pilferage.
7. *The aid should provide for a test of learning progress.* The trainee should be able to recognize the actual article after seeing the aid and distinguish it from among several similar abjects.
8. *The aid should be inexpensive, rugged, and safe.*
9. *The aid must be properly designed.* Glare must be controlled, height of mounting should be realistic, and all matters of set-up should be arranged properly before the presentation.
10. *The aid must have "appeal."* It must be attractive in color, movement, or form, but not at the expense of clarity and accuracy. For example, the improper use of a bright color might direct the attention of the trainee away from significant parts

There is general acceptance of the notion that training aids add to the effectiveness of training, and there is experimental evidence to lend support to this belief. There is, however, also a danger in the use of training aids, the danger being that the instructor will become "training-aid happy" and rely on them to the detriment of the effectiveness of the training. In other words they should be used judiciously, and, where used, they should be appropriate to the purpose.[62]

Conferences Conferences are used particularly in supervisory and management training, but are also used for other types of training, such as sales training. This method is most appropriate for the purposes of: (1) developing the problem-solving and decision-making facilities of personnel; (2) presenting new and sometimes complicated material; and (3) modifying attitudes.

The particular purpose of a conference will serve as the basis for the manner in which it is carried out. If the purpose is that of developing the problem-solving and decision-making skills of the participants, the conference leader must adroitly facilitate the participation of the individuals, but at the same time prevent the conference from straying too far from the objective. In case the conference is to result in an actual decision to a real problem, as well as to serve training purposes, management must be willing to accept whatever decision is made by the group. Otherwise, the problem in question should not be posed as one for a conference decision. If the conference is intended primarily as an avenue for presenting information (such as a new policy that is being adopted, or a new procedure that is to be implemented), the conference leader typically would play a more active role but encourage questions and discussion on the part of those present. When the conference is directed more toward modification of attitudes, its "direction" may be extremely

---

[62] Those interested in the preparation of training materials are referred to *Guide to the Preparation of Training Materials,* Bureau of Employment Security, U.S. Department of Labor (Available from the Superintendent of Documents, Washington, D.C., price 50 cents).

limited; in fact, in the case of some training programs (such as sensitivity training) there may be no direction. The lack of structure or direction usually permits a much higher degree of participation and interaction than in the case of other types of conferences.

In terms of psychological principles involved, it is probable that the most important one is that of active participation. In addition, however, the conference provides the opportunity for "reinforcement" by the trainer of such participation, as pointed out by McGehee and Thayer.[63] However, the reinforcement should be for the participation as such, and should not consist of verbal rewards or punishments for the *nature* of the participation—otherwise the conference leader loses his neutrality and is taking sides.

Sensitivity Training Sensitivity training stems from the theories of group dynamics. While the recognition of the phenomenon of group interaction is nothing new, Lewin[64] systematically investigated this phenomenon over a period of years, and focused attention on this process. He reported, for example, that group discussion was markedly more effective in bringing about change of attitudes than was a lecture presentation. Taking a cue from this, training programs have been developed that, in effect, provide for group discussion of problems. This method is variously referred to as sensitivity training, action research, t-group training, and d-group training.

In sensitivity training it is usually the practice to have the training group brought together in a place away from their jobs for a period of several days. During this time they are under the very general direction of a training leader, but his role is a very nominal one, typically that of simply setting the stage and observing. He may or may not assign some topic for discussion. Basically, however, the group is on its own. As it tries to organize itself and grapple with discussion problems, the interaction of the members comes into play, with the associated human reactions to the successes, frustrations, failures, personal differences, irritations, and jealousies that usually are manifest. Out of the welter of this unstructured, sometimes aimless, process, the participants are supposed to develop greater understanding of the behavior of others, and sensitivity to their attitudes (hence the name "sensitivity training").

The effectiveness of this method would be reflected by any ultimate desirable change on the part of the participants on their jobs. At the present time there is very little evidence regarding this. For what interest the self-testimonials on the part of participants might be, opinions of participants on certain aspects of this type of training will be summarized. The training program in question was carried out by a large petroleum

---

63 McGehee and Thayer, *op. cit.*

64 K. Lewin, *Field Theory in Social Science* (New York: Harper & Row, Publishers, 1951).

refining company.[65] The participants were formed into groups of 8 for a period of 3 days. At the beginning and end the participants completed a "perceptionnaire" in which they in effect rated themselves on each of several behavior areas, including the following:

Behavior Area	*Mean Self-rating* *Before*	*After*
Problem solving	5.9	6.4
Problem diagnosis	5.7	6.2
Consideration	6.6	6 8

As an additional indication of the reactions of the participants, 86 were brought together 6 months later, and were asked to give their opinions about their experiences in the training program. These were in the form of ratings of the impact of the training in terms of three questions; the results are given in Fig. 10.14. The somewhat higher mean values for the first (and second) questions suggest that the participants felt that the

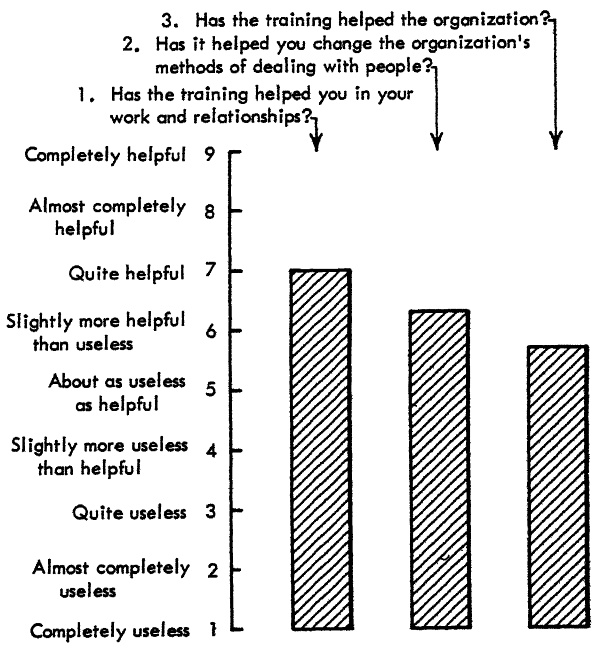


Fig. 10.14. Mean ratings of answers to three questions regarding the impact of sensitivity training. These ratings were obtained from 86 men who had been through such training six months before. (From *An Action Research Program for Organization Improvement, op. cit.*, p. 31.)

[65] *An Action Research Program for Organization Improvement* (Ann Arbor, Mich.: The Foundation for Research on Human Behavior, 1960).

training had helped *them* more than the *organization*. In both of these sets of data, the responses indicate somewhat favorable reactions to sensitivity training on the part of the participants. While these opinions are suggestive of benefits from the program, they need to be accepted with some reservations, since they may be influenced in part by a tendency (perhaps entirely unintentional) to report what is "expected."

The actual effect of such training in terms of the behavior of the individuals when they get back on the job actually is unknown. As pointed out by Shepard,[66] while some really remarkable problem solving occurred during the training itself, it is too early to say what the long-run consequences will be. As very appropriately pointed out by McGehee and Thayer,[67] a primary question relates to that of reinforcement of behavior. While modified (desirable) behavior may be reinforced by the fellow participants during the training sessions, there is a question as to whether such behavior will also receive reinforcement back on the job.

Case Study The case study method is one in which an actual or hypothetical problem is presented to the group for group solution. It is another form of conference. Typically, the case in question is presented in written form in advance of the meeting. While the nature of the cases can range over a wide gamut of topics, the presentation states the circumstances of the case, including a detailed write-up of pertinent material, and poses the problem as it is to be considered by the group. An example of the introduction to an illustrative case is given below.[68]

The Questionable Hiring Date

THE INCIDENT

On August 1, 1953, the Employment Manager sent the following memo to Arthur E. Holz, employee No. 210096:

> We regret that, owing to a clerical error, your company seniority was incorrectly noted as dating from December 3, 1951. The error has now been corrected. This is to inform you that your company seniority dates from March 17, 1952.

This memo made Mr. Holz angry. In filing a complaint about it, he said to his Shop Steward: "Some correction, brother! It was all right before. But now, they've really fouled things up."

ASSIGNMENT TO GROUP MEMBERS

You are the impartial arbitrator to whom this case has been referred after failure of the parties to settle it themselves. You will be expected to render an Award and to outline your Discussion (reasoning).

---

66 H. Shepard, "An Action Research Approach to Organizational Development," *The Management Record*, 22 (No. 6, June, 1960), 26–30.

67 McGehee and Thayer, *op. cit.*, p. 206.

68 Reproduced with permission from: P. and F. Pigors, *Manual for Group Members: The Incident Process, Practical Supervision Problems, Series I* (Washington, D.C.: The Bureau of National Affairs, Inc., 1955). This manual includes several cases dealing with supervisory problems. Another source book of cases is *Cases on Human Relations in Management*, by R. P. Calhoon, E. W. Noland, and A. W. Whitehall, Jr. (New York: McGraw-Hill Book Company, 1958).

The above represents only the introduction to the case; it is supplemented by various attachments, such as the applicable contract provisions, and a presentation of pertinent background information on some of the factors related to probationary status and seniority.

The team leader in the case study method is expected to guide and to stimulate independent thinking that is also cooperative. The case study method has been characterized as being made up of the following phases: [69]

1. Studying the incident (the case).
2. Getting the facts.
3. Determining what needs to be decided.
4. Deciding what should be done about it.
5. Evaluating behavior (connecting the case to other situations, and thinking ahead toward possible measures that might be useful in that kind of situation, and that might apply in the future).

The basic objective of the team discussion of cases has been stated by Pigors and Pigors [70] as *learning by doing*. It has been suggested that the more important contributions of the case method include [71] the following: discouraging snap judgments about people; looking for "the" answer; illustrating how the same set of events can be perceived differently; destroying smug generalizations; training in discussion and interaction with others; and keeping thinking in a practical setting.

It has been proposed that the case study method is especially appropriate under the following circumstances: [72] when employees need to be trained to identify and analyze complex problems and to frame their own decisions; when employees need to be exposed to a variety of approaches, interpretations, and personalities (that there are, in other words, very few pat answers to business problems); when the personnel are sufficiently sophisticated to draw principles from actual cases, and formulate solutions to problems themselves; and when a challenge is needed for overconfident individuals. On the other hand, the procedure should be avoided with "beginners" or persons with lack of maturity, or when internal jealousies and tensions make people reluctant to air their opinions and ideas freely.

Role Playing Role playing has been described as a technique of creating a life situation, usually one involving conflicts between people, and then having persons in a group play the part of specific personal-

---

[69] P. Pigors, C. A. Myers, and F. T. Malm, *Readings in Personnel Administration* (New York: McGraw-Hill Book Company, 1959), pp. 253–254.

[70] Pigors and Pigors, *ibid.*, p. 306.

[71] N. R. F. Maier, A. R. Solen, and A. A. Maier, *Supervisory and Executive Development: A Manual for Role Playing* (New York: John Wiley & Sons, Inc., 1957), p. 2.

[72] Proctor and Thornton, *op. cit.*, p. 81.

ities.[73] In industry it is used primarily as a technique for modifying attitudes, and for developing skills in interpersonal relationships. It is, therefore, particularly suitable for training individuals on those jobs which require such skills, as in the case of supervisors, managerial personnel, salesmen, etc. In role playing, each participant plays the "part" (role) of someone in a simulated situation. Typically, the (hypothetical) situation is structured by setting forth the "facts" of the situation, such as the job situation in question, the events that led up to the "present" situation, and other relevant information. It is in these respects much like the case study. In role playing, however, individuals are then designated to play the roles of the persons in the "case." Some of the cases described by Maier [74] include such problems as unscheduled coffee breaks, the use of office phones, a bottleneck in an assembly operation, and the allocation of overtime. Once the case has been stated and the individuals assigned to their roles, they "play" those roles as though the situation were real. Thus, a supervisor might play the role of a subordinate, a salesman the role of a customer, and a nurse's aide the role of a hospital patient. Having to put one's self in someone else's place, and play the part of that person, generally increases one's empathy for the other person and an understanding of his behavior. The process also helps a participant to appreciate the fact that the behavior of others is the consequence not only of their personalities but also of the situation, and helps him to identify and recognize some of his own faults, or behaviors, that might rub other people the wrong way.

Management Games A management game has been described as a dynamic training exercise utilizing a model of a business situation.[75] In these games, the trainees (such as executives) make the same kinds of operating and policy decisions as are required in real life. Usually several participants comprise a "team," with various teams being placed in competition with each other. Typically each team represents a "company," and is presented with a statement of the problem about which the game revolves. Some typical problems, presented by Greene and Sisson,[76] are: materials inventory; personnel assignment; retailing operations; production scheduling; industrial sales; top management decisions; and market negotiation. Whatever the problem, the team is provided with appropriate information; depending on the problem, this information can cover such factors as assets, inventories, labor costs, storage costs, demand curves (the demand for a product as it varies with

---

[73] N. R. F. Maier, *Principles of Human Relations* (New York: John Wiley & Sons, Inc., 1956), p. 87.

[74] Maier, Solem, and Maier, *op. cit.*

[75] J. M. Kibbee, C. I. Craft, and B. Nanus, *Management Games: A New Technique for Executive Development* (New York: Reinhold Publishing Corp., 1961), p. 3.

[76] J. R. Greene and R. L. Sisson, *Dynamic Management Decision Games* (New York: John Wiley & Sons, Inc., 1959).

price), and interest charges. The team then organizes itself (such as selecting a president and other officials), and proceeds to make decisions about its policies and operations. The effects of these decisions typically can be computed in quantitative terms, such as profits, this being done either by computers or by hand. At the end it is possible to determine what team "won." While the game may take hours, days, or even months, it actually represents a longer time span; thus, the theoretical total time is compressed into the time spent on the game itself. The game is then followed by a critique in which the actions of the teams are analyzed.

This method of training provides active participation in a somewhat life-like situation, with the opportunity for feedback of the consequences of one's (actually the team's) decisions. It has been suggested that a player might benefit in some of the following ways from participation in such a game: [77] the player should learn which key factors to observe in the actual on-the-job situation; his attention is focused on established policies or strategies and on longer-range planning, rather than on "putting out fires"; he learns to make better use of decision-assisting tools such as financial statements and statistical inventory control; the game illustrates the value of analytic techniques such as the use of mathematical models to arrive at "optimum" solutions; and feedback allows for actions taken in one "period" to affect future conditions and results.

There is as yet little factual evidence of the long-range effects of management games as a training procedure. Subjective evaluations of participants and observers, however, generally are very favorable. Perhaps more significant, however, is the rational argument based on transfer-of-learning concepts. To the extent that the games do parallel the problems and situations of the real world, it would be expected that some of the "learning" would rub off on the participants, and would, therefore, have a carry-over effect to their jobs.

A brief comment should be added about the development and availability of games. A number of games have been developed, many of which require the use of computers. A directory of such games is given by Kibbee, Craft, and Nanus.[78] To encourage the use of games where computers may not be available, Greene and Sisson [79] present seven non-computer games, and also set forth instructions for constructing such games.

Other Training Methods There are numerous other varieties of training methods that have been used in various situations. A few of these will simply be mentioned briefly. These include planned job rotation, understudy training, on-the-job "coaching" (guidance to the subordinate by the superior), and training carried out outside the organization, such

---

77 Adapted from Greene and Sisson, *op. cit.*, pp. 3–4.
78 Kibbee, Craft, and Nanus, *op. cit.*, pp. 315–336.
79 Greene and Sisson, *op. cit.*

as by sending personnel to special college or university workshops and programs.

A technique used by the armed forces with great success during World War II was the tachistoscopic airplane and ship recognition training. This approach, founded upon the basic work of Renshaw,[80] involves presentation of stimuli for a very brief period (often 1/10 of a second or less). By the use of this method, subjects have been taught to see and identify distinguishing features of objects or forms when the "flash" presentations are reduced to an exceedingly small fraction of a second. In fact, Renshaw found that when trained by this method ordinary people rapidly achieve a degree of skill heretofore considered genius. Although this method of training has apparently not been used extensively in industry, it would seem to offer real possibilities for certain types of training, particularly on such jobs as inspection for appearance. This method does serve as the basis of several developmental reading courses, some of which are used in industry for increasing the reading rate of personnel.

THE EVALUATION OF TRAINING

It is probable that most organizations assume that their training programs are achieving their intended objectives. Such faith, however, may sometimes be unwarranted. If an organization really wants to know whether its training program is accomplishing its purposes, it must go through a systematic evaluation process. There usually are few short-cuts in this.

In general, such evaluation is directed toward determining how effective the training program is in helping *groups* of employees acquire the desired skills, knowledges, and attitudes. While a trainer needs to be perceptive of the progress of individuals during training, the program as such should be evaluated in terms of the progress of the group generally.

It should be pointed out that while the evaluation of training usually is considered in the context of "training" versus "not training," it is also appropriate to evaluate the relative effectiveness of different methods of training.

Basis of Evaluation of Training The evaluation of training involves the use of an appropriate criterion. In the selection of the criterion, the same considerations should be used that are pertinent to the selection of criteria for other purposes. As discussed in Chapter 2, these considerations include *relevance, reliability,* and *freedom from contamination.* Many of the same types of criteria used in test validation are also appropriate in training evaluation, as, for example, quantity and quality of work. time

[80] S. Renshaw, "The Visual Perception and Reproduction of Forms by Tachistoscopic Methods," *Journal of Psychology,* 20 (1945), 217–232.

to perform some operation, performance tests, and in some cases ratings. In training evaluation, however, some of these criteria may be converted into other forms, such as time to reach a given level of performance, or percentage of improvement during training.

Levels of Evaluation The evaluation of training can be carried out at different levels. As indicated by MacKinney,[81] for example, training evaluation can be based on *subjective* or *objective* criteria. In terms of the methods used, it can be considered as being either *informal* or *formal*. The evaluation would, of course, be more adequate if it is carried out in a formal, systematic manner, and with objective criteria (unless there is a justifiable basis for using subjective criteria). With specific reference to the evaluation of supervisory training, the following levels of evaluation have been proposed by Lindbom and Osterberg: [82]

1. Supervisors' classroom behavior. (This is based on the assumption that performance in training is related to performance on the job.)
2. Supervisors' behavior on the job.
3. Employees' behavior on the job. (One can use such measures as productivity, productivity ratio, quality, and morale.)

While each of these levels has its place, the third level, employees' performance on the job, usually would be the most meaningful since the objective of supervisory training usually is that of training supervisors to increase the effectiveness of the operations for which they are responsible.

Basic Methods of Evaluation Three basic approaches to training evaluation have been characterized by MacKinney.[83] The first of these is *controlled experimentation.* This procedure requires the use of two groups of employees, namely, a "training" group and a "control" group. Criterion measures are obtained for both groups *before* training is started. The training group is then given the training, while the control group is not (the group continues to perform the job without training). After the first group is trained, criterion measures are obtained for both groups. A subsequent comparison is then made of the improvement in both groups to determine if the training group improves significantly more than the control group. The second method involves the use of a single training group with criterion measures being obtained *before and after* the training. The third method also involves a single training group but with criterion measures being obtained *after* the training (but not before).

Of these, the first method (controlled experimentation) is suggested by MacKinney as the "only" procedure that really provides a solid basis for evaluation of the training, and expresses the opinion that training

---

81 A. C. MacKinney, "Progressive Levels in the Evaluation of Training Programs," *Personnel,* 34 (No. 3, November–December, 1957), 72–78.

82 T. R. Lindbom and W. Osterberg, "Evaluating the Results of Supervisory Training," *Personnel,* 31 (1954), 224–228.

83 MacKinney, *op. cit.*

people should not waste time and money on anything short of such a procedure. The second method (with "before" and "after" criterion measures) does not provide the basis for knowing how much improvement would have occurred *without* training. The third method (with only an "after" measure) obviously does not provide any basis for knowing how much improvement actually occurred during training.

While it is sometimes difficult (in terms of practical considerations) to pursue the controlled experimentation procedure, this procedure demonstrates the principle of applying scientific, experimental methods to the personnel problems of industrial organizations.

Examples of Training Evaluation For illustrative purposes, a few examples of training evaluations will be presented.

Record press operation. In a company manufacturing music records, a training program was initiated for training new press operators. Previously, new operators had been started on the job without formal training. A comparison of the production of one group of operators who went through the training program versus another group who did not is reported by Lawshe.[84] This comparison, based on the percentage who met specified production standards, is shown in Fig. 10.15 for both the first 4 weeks and the first 8 weeks. The results in this case show a significant difference between the 2 groups for both time periods.

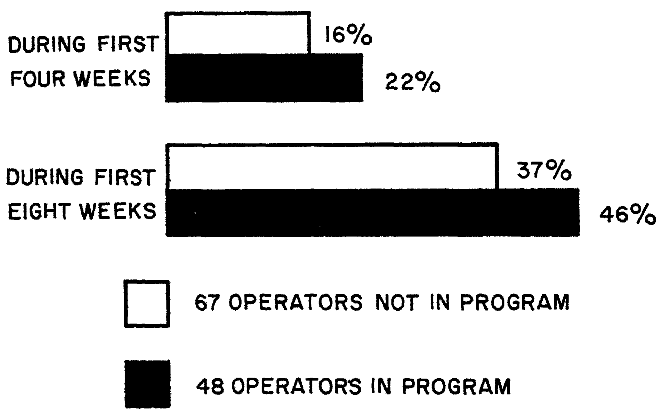


Fig. 10.15. Effect of an "induction" program upon per cent of employees engaged in a record press operation who met production standards after four and eight weeks, respectively. (Figs. 10.15 to 10.17, inclusive, courtesy Factory Management and Maintenance, from C. H. Lawshe, Jr., "Eight Ways to Check the Value of a Training Program," Factory Management and Maintenance, 1945, 103, 117–120.)

[84] C. H. Lawshe, Jr., "Eight Ways to Check the Value of a Training Program," *Factory Management and Maintenance*, **105** (1945), 117–120.

Tungsten-disc cutters. Previous mention was made of the tungsten-disc cutting operation studied by Lindahl.[85] A training program was developed that provided for feedback to the operators of their own pedal action, as recorded graphically in the manner illustrated in Fig. 10.7. One criterion that was used consisted of the breakage rate of the grinding wheels that were used (the grinding wheels were easily broken by improper pedal action). Fig. 10.16 shows the decrease during training of

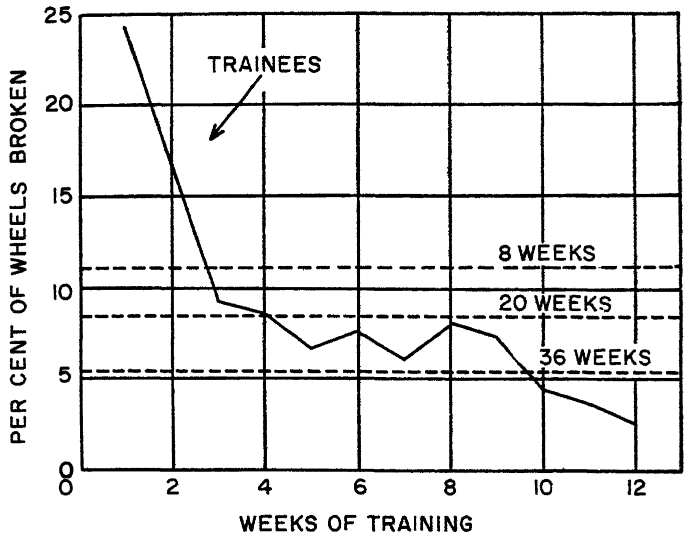


Fig. 10.16. Decrease in abrasive wheel usage of trainee group in comparison with present employees of varying amounts of experience.

the wheels broken during the operation. That figure also shows the corresponding breakage rate of a previous group of untrained operators at 8, 20, and 36 weeks on the job. It can be seen that the trained group achieved in 10 weeks a rate that previously was achieved only after 36 weeks.

Labor turnover. The expense of training new employees usually is a significant aspect of production costs. The voluntary termination of personnel sometimes is brought about by unfavorable attitudes on the part of employees. Training programs sometimes can aid in developing greater interest on the part of employees, and in preventing employees from becoming unduly discouraged about their progress during training.

[85] Lindahl, *op. cit.*

In one company a vestibule school was set up to train new personnel.[86] After the program had been in operation a while, a comparison was made of the terminations during the first 30 days of 70 personnel who went through the training program and of 70 who previously had been placed immediately on the job for on-the-job training. The comparison, shown in Fig. 10.17, shows a distinct difference in favor of those who attended the vestibule training school.

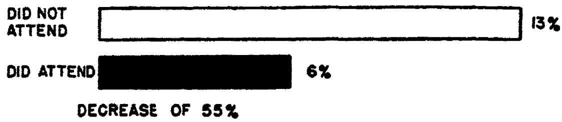


Fig. 10.17. Comparison of terminations before 30 days of employment of employees attending vestibule school and employees broken in "on the floor."

[86] Lawshe, *op. cit.*

11

The Measurement of Attitudes and Morale

Management has a growing interest in employee attitudes, opinions, and morale. This interest is reflected in the fact that many companies conduct attitude and opinion surveys, provide "human relations" training for supervisors, sponsor recreation programs for employees, publish employee newspapers, and take other actions that are intended to create favorable attitudes on the part of employees.

This concern on the part of management probably can be attributed in part to a general trend toward greater recognition of social responsibilities by industry. In part it can probably be attributed to the belief on the part of management that employees with favorable attitudes toward their company generally are "better" employees in some respect, such as being more productive, or having lower turnover rates. There is some evidence to support this belief, but a discussion of this will be deferred until later.

Before considering some of the factors that affect attitudes and morale, let us discuss their measurement. Even before doing this, however, we should crystallize some of the concepts that we will be discussing.

THE MEANING OF TERMS

Attitudes The term *attitude* has various connotations; we will use it in the sense defined by Maier.[1] He defines an *attitude* as a kind of mental set. It represents a predisposition to form certain opinions. To put it another way, it is a frame of reference that influences the individual's views or opinions on various topics, and that influences his behavior.

Attitudes are formed through experience, which means that they are learned. Once a person has developed a particular attitude, it may be difficult for him to determine how he acquired it. In fact, an individual may not be consciously aware of his own fundamental attitudes. Whether a person's attitudes are based on rational considerations and factual information, or whether they have a strong emotional bias, has little bearing on the *effect* of the attitudes on the person's thinking or behavior. In either case the factor that affects behavior is the *attitude, not* the consideration of whether it is or is not a rational attitude.

Although the attitudes of people tend to be relatively stable, they can be modified, at least to some degree. Since they can be changed, it is more fruitful to measure them than would be the case if they were firmly fixed, once and for all.

Opinions An *opinion*, on the other hand, is an expression of an evaluative judgment or point of view regarding a specific topic or subject. Thus, an attitude is somewhat generalized (such as liking or not liking a person's supervisor), whereas an opinion typically is an interpretation regarding a specific matter (such as saying that the boss plays favorites in granting overtime). Opinions, however, typically are influenced by the more generalized attitude. The facts or observations within an individual's experiences are *interpreted* in the light of his *attitudes*. Thus, if a foreman calls the attention of his work group to the fact that some of the safety rules have been violated, one person (who has an "unfavorable" attitude toward the foreman) might later express the opinion to one of his buddies that the foreman is "—just picking on us"; another person (who has a "favorable" attitude toward the foreman) might later express the opinion that the foreman is simply "—trying to keep us from getting our fingers cut off."

Morale Guion,[2] after considering several definitions, defines morale as *"the extent to which an individual's needs are satisfied and the extent to which the individual perceives that satisfaction as stemming from his total job situation."* While this definition covers the details of morale, it does not consider the *group* reaction aspect which would seem to be essential to the concept of morale. Whereas *attitudes* and *opinions* are

[1] N. R. F. Maier, *Psychology in Industry*, 2nd ed. (Boston: Houghton Mifflin Company, 1955).

[2] R. M. Guion, "Some Definitions of Morale," *Personnel Psychology*, 11 (1958), 59–61.

characteristic of *individuals,* the term *morale* generally has an implication of *group* reactions. In other words, the morale of a group depends on the interactions among individuals in the group, such as a football team, a military unit, or a work group. It is akin to the common notion of team spirit.

MEASUREMENT OF ATTITUDES AND OPINIONS

Attitudes, opinions, and morale cannot be weighed on a scale, like a pound of butter, nor measured with a rule, like a strip of carpet. Yet, though they are intangible, they can be measured. Thorndike long ago pointed out that whatever exists, exists in some amount, and whatever exists in some amount can be measured (though this is sometimes admittedly difficult). Certain differences in employee attitudes do exist, and, further, they can have implications for the effectiveness of the individuals within an organization, and for the organization as a whole. The morale of the people within a group, for example, can make or break a business enterprise or a military campaign.

There are various ways in which it is possible for supervisors and managers to get some inkling of the attitudes and opinions of individuals and of the morale of a group. Some of these methods are very informal and unsystematic, such as interpreting chance remarks of individuals, the behavior of individuals and work groups, and the reported "feelings" of supervisors. A sensitive foreman, for example, frequently can get a "feel" with respect to the general reactions of those in his work group even though he cannot pin such reactions down specifically.

Another method of getting some inkling about employee feelings is by analyzing changes in such factors as turnover rate, absenteeism, and production level. These clues, however, are at best only indirect, and they can also be influenced by factors *other* than the reactions of employees.

For many purposes it is desirable to obtain some more adequate, and direct, measure of attitudes, opinions, and morale. Various methods have been developed for doing this, some of these consisting of questionnaires which are filled out anonymously by employees. Some of these methods will be described and illustrated.

There are two general types of questionnaires in common usage. One is sometimes called an *attitude scale,* and the other an *opinion survey.*

ATTITUDE SCALES

The specific questions of typical attitude scales may call for an individual to give his "opinion" about some aspect of the company, but these "opinions," collectively, are used to obtain some general attitude level for the (anonymous) individual represented by each questionnaire.

The average of the attitude scores for the employees within a group (job, department, company, or whatever), in turn, is sometimes used as an index of "morale" of the group. There are various types of attitude scales, but they all serve essentially the same purposes.

Thurstone Type of Scale One type of attitude scale is the Thurstone scale. It can be used to determine the general attitude of employees toward their company. An example is shown in Table 11.1. This scale is

Table 11.1

STATEMENTS USED IN BERGEN'S SCALE FOR MEASUREMENT OF ATTITUDE OF EMPLOYEES TOWARD THEIR COMPANY

	Scale value
I am made to feel that I am really a part of this organization	9.72
I can feel reasonably sure of holding my job as long as I do good work	8.33
I can usually find out how I stand with my boss	7.00
On the whole, the company treats us about as well as we deserve	6.60
I think training in better ways of doing the job should be given to all employees of the company	4.72
I have never understood just what the company personnel policy is	4.06
In my job I don't get any chance to use my experience	3.18
I can never find out how I stand with my boss	2.77
A large number of the employees would leave here if they could get as good jobs elsewhere	1.67
I think the company's policy is to pay employees just as little as it can get away with	.80

taken from an article by Bergen.[3] The theory of such scales and detailed instructions for their construction go back to the early work of Thurstone and Chave.[4] This same procedure was mentioned in Chapter 9 on "Performance Appraisal," since it also lends itself to use in the development of weighted check list performance rating systems.

In the development of attitude scales by this method, the first step is to write out a large number of statements, perhaps a hundred or more, each of which expresses a viewpoint of some kind toward the company. An effort should be made to have these statements express all possible viewpoints from extremely favorable to extremely unfavorable. Each of these statements is typed on a separate slip of paper and a judge is asked to place each statement in one of several piles (usually 7, 9, or 11), ranging from statements judged to express the *least* favorable viewpoints (placed in pile 1), to statements judged to express the *most* favorable viewpoints (placed in the top pile, 7, 9, or 11). Statements judged to express varying

[3] H. B. Bergen, "Finding Out What Employees Are Thinking," *The Conference Board Management Record*, April, 1939.

[4] L. L. Thurstone and E. J. Chave, *The Measurement of Attitude* (Chicago: University of Chicago Press, 1929).

degrees of favorableness in between these extremes are placed in the piles that are judged best to characterize their relative degrees of favorableness. A record is kept of the category in which each judge classifies each of the many statements.

Many judges are used in this process, sometimes as many as 100 or more. It should be emphasized that these judges are assisting in the construction of the scale. They are not having their own attitudes measured. To measure attitudes with the scale is not possible until the scale has been constructed, and the allocation of statements to the several piles is a part of the process of constructing the scale.

The purpose of the allocation is to determine the scale values of the various statements. If all judges tend to place a statement in piles toward the favorable end of the continuum, we may safely conclude that that statement expresses a favorable attitude toward the company. If a statement is generally placed by the judges in piles toward the unfavorable end of the series, we may likewise conclude that an unfavorable attitude is expressed by that particular statement. The piles are numbered from the unfavorable to the favorable end of the continuum. The number of times each statement is placed in each pile is determined, and a computation made to determine the average location of the statement by the judges. From this calculation, the scale value of the statement is determined, as shown in the example in Table 11.1. It should be added that an analysis is also made of the consistency of the judgments for each statement. Statements that are placed by all judges in one or in a limited number of categories are the statements that have the greatest degree of stability. Statements that are scattered by the judges over several categories are eliminated. By starting with many more statements than need be retained for the final scale, it is usually possible to identify 10 to 15 statements that are spread over the entire range, and that have been consistently judged by the judges. The statements so selected, together with their scale values as shown in Table 11.1, comprise the final material for the attitude scale.

One might think that the attitudes or feelings of the judges who are used in the construction of the scale would have an effect on the scale values obtained. In other words, it might be felt that one set of scale values for a series of statements might be obtained if the judges were, in general, *favorable* toward the company, while a different set of values might be obtained if the judges in general were unfavorable or indifferent toward the company. This possibility has been subjected to experimental test, and it has been found that the attitudes of the judges do not significantly affect the scale values obtained from them. This fact increases the possibilities of the use for such scales in industry because it often happens that the persons who are most conveniently available for use as judges in constructing a scale may be more favorably disposed toward

the company than certain groups of employees with whom the scale is to be used after it is constructed.

In the practical administration of an attitude scale, statements are printed on a sheet in random order, without the scale values appearing in Table 11.1. Each employee is given one of these sheets and is requested to check all statements that he agrees with or believes to be true. The sheets are then turned in without being signed. The attitude of an employee toward the company usually is defined as the average or median scale value of the statements he has checked. If the average is to be used, an employee checking statements 1, 3, and 5 of those shown in Table 11.1 would have an attitude score of:

$$\frac{9.72 + 7.00 + 4.72}{3} \text{ or } 7.15$$

On a scale of ten (ten being the most favorable end and zero the least favorable end) an attitude of 7.15 would be one somewhat toward the favorable end of the scale. On the other hand, an employee checking statements 7, 8, and 10 would have an attitude score represented by:

$$\frac{3.18 + 2.77 + .80}{3} \text{ or } 2.25$$

This would be a much less favorable attitude toward the company than the one described above.

Mention was made in Chapter 9 of a possible fallacy in the use of the median in merit rating, as pointed out by Jurgensen.[5] This same fallacy also can occur in the use of the Thurstone type of scale in attitude measurement. The solution is described in Chapter 9 (see p. 237).

The attitude scale reproduced in Table 11.1 is by no means the only set of statements that might be constructed to fulfill the requirements of a suitable scale. Uhrbrock,[6] in a comprehensive study of the attitudes of 3,934 factory workers, 96 clerical workers, and 400 foremen, used a different scale, but one constructed as explained above to measure the general attitude of employees toward the company. Some typical statements, with their scale values from Uhrbrock's scale, are reproduced in Table 11.2.

The fact that scales such as those illustrated in Tables 11.1 and 11.2 may be used independently to measure the attitude of employees toward their company illustrates several points concerning the use of such scales. It will be noticed that the statements comprising the two scales are not the same, yet both scales measure essentially the same thing. Thus, the

[5] C. E. Jurgensen, "A Fallacy in the Use of Median Scale Values in Employee Check List," *Journal of Applied Psychology*, 33 (1949), 56–58.

[6] R. S. Uhrbrock, "Attitudes of 4,430 Employees," *Journal of Social Psychology*, 5 (1934), 365–377.

use of such scales does not require including any specific statements that, from an industrial relations viewpoint, might seem undesirable. A scale can be "tailor-made" for a given plant and deliberately kept free of statements that may seem to "pack dynamite" without sacrificing the validity of the scale. The use of different statements in scales measuring the same attitude also facilitates checking results by a repeat test in order

Table 11.2

STATEMENTS USED IN UHRBROCK'S SCALE FOR MEASURING ATTITUDE OF EMPLOYEES TOWARD THEIR COMPANY

Statement	*Scale value*
I think this company treats its employees better than any other company does	10.4
If I had to do it over again I'd still work for this company	9.5
They don't play favorites in this company	9.3
A man can get ahead in this company if he tries	8.9
I have as much confidence in the company physician as I do in my own doctor	8.7
The company is sincere in wanting to know what its employees think about it	8.5
A wage incentive plan offers a just reward for the faster worker	7.9
On the whole the company treats us about as well as we deserve	7.4
I think a man should go to the hospital for even a scratch, as it may stop blood poisoning	6.3
I believe accidents will happen no matter what you do about them	5.4
The workers put as much over on the company as the company puts over on them	5.1
The company does too much welfare work	4.4
Soldiering on the job is increasing	4.1
I do not think applicants for employment are treated courteously	3.6
I believe many good suggestions are killed by the bosses	3.2
My boss gives all the breaks to his lodge and church friends	2.9
I think the company goes outside to fill good jobs instead of promoting men who are here	2.5
You've got to have "pull" with certain people around here to get ahead	2.1
In the long run this company will "put it over" on you	1.5
The pay in this company is terrible	1.0
An honest man fails in this company	0.8

to be sure of conclusions reached and to measure the effectiveness of systematic company efforts to improve employee morale.

Likert Type of Scale A number of years ago Likert[7] developed a method of measuring attitudes that is somewhat simpler than the Thurstone method, and avoids the necessity of using unfavorable statements. Managements often object to the use of such statements. Likert's method, moreover, does not require the use of judges in scaling the statements. While a number of different procedures were tried and compared, the simplest method described by Likert was found to give results that

[7] R. Likert, "A Technique for the Measurement of Attitudes," *Archives of Psychology*, 1932, No. 140, 55 pp.

correlated very highly with more complex methods. This method consists in printing after each statement five degrees of approval, and asking the person taking the scale to check one of the five. A typical statement from Likert's Internationalism Scale is as follows:

> The United States should have the largest military and naval air fleets in the world. Strongly approve ( ) Approve ( ) Undecided ( ) Disapprove ( ) Strongly disapprove ( )

These responses are scored from 1 (strongly approve) to 5 (strongly disapprove). By using approximately twenty-five such statements, each followed by five degrees of approval, the sum of the scores of the responses checked constitutes a measure of one's attitude toward the object of the scale. Since this method can be followed by using only statements that vary between indifference (the central point on an attitude continuum) and the favorable end of the scale, the use of unfavorable statements in the scale itself is avoided. Likert has shown that this method of measuring attitude correlates approximately .80 with the more tedious Thurstone method.

Attitude scales are not perfect instruments for the registering of employee feelings, but they are considerably better than guessing or relying on chance (and often biased) individual reports.

OPINION SURVEYS

Attitude scales of the type described in the preceding section help to measure the attitudes of individuals; by summarizing data for all employees within a group, such scales can be used to quantify the "morale" of employee groups. Comparisons then can be made from one department to another. However, although such scales can be useful in indicating the relative level of morale of employee groups, they do not enable a management to identify specific factors that may be sources of employee unrest or dissatisfaction. This information can be obtained by the use of questionnaires that provide for giving opinions about specific matters, such as working conditions, company policies, or food in the cafeteria. An example of a questionnaire used in an employee survey of this type conducted by the Victor Adding Machine Company of Chicago is shown in Fig. 11.1. The questionnaires were distributed by mail to the employees' homes, after the employees had been informed that the questionnaires were to be distributed. They were mailed anonymously by the employees to a tabulating agency.

The usual practice in opinion questionnaires is that of obtaining a single response to each item or question such as the "yes" or "no" as illustrated in Fig. 11.1, such responses indicating the degree of satisfaction with, or an opinion about, some specific aspect of the total work situation.

QUESTIONNAIRE

1. This questionnaire is being answered by: Man ☐ Woman ☐
2. Are you— Married ☐ Single ☐
3. Are you paid by the week or by the hour? Paid by the week ☐ Paid by the hour ☐
4. Do you feel you would rather be doing some other type of work? Yes ☐ No ☐
   If yes, have you discussed it with the Personnel Office? Yes ☐ No ☐
5. How do you regard safety conditions within the plant generally? Good ☐ Not so good ☐

What is your opinion of your foreman or department head?

6. Does he "know his stuff"? Yes ☐ No ☐
7. Does he play favorites? Yes ☐ No ☐
8. Does he keep you busy? Yes ☐ No ☐
9. Does he keep his promises? Yes ☐ No ☐
10. Does he pass the buck? Yes ☐ No ☐
11. Does he welcome suggestions? Yes ☐ No ☐
12. Is he a good teacher? Yes ☐ No ☐
13. Do the workers know more than he does? Yes ☐ No ☐
14. Does he set a good example? Yes ☐ No ☐
15. Do you think you are in need of more training? Yes ☐ No ☐
16. Which of the following ways of pay would you prefer? Present way ☐ Incentive or piece work ☐

Do you feel you understand the following provisions of the Company Security Fund?

17. The insurance provisions Yes ☐ No ☐
18. How the Security Fund shares are figured Yes ☐ No ☐
19. How VAMCO determines its contribution to the Fund each year Yes ☐ No ☐
20. How the Security Fund money is invested Yes ☐ No ☐
21. How your Fund account is closed in event of termination, death or retirement Yes ☐ No ☐

22. Do you feel that you are receiving considerate treatment here? Yes ☐ No ☐
    If not, why____________________

23. Do you feel top management is interested in the employes? Yes ☐ No ☐
24. Have you ever recommended Victor as a place to work to a friend? Yes ☐ No ☐
25. Would you like more news broadcasts given over the P. A.? Yes ☐ No ☐
26. Are you interested in Company athletic activities? Yes ☐ No ☐
27. Are you making suggestions to the Suggestion System? Yes ☐ No ☐
28. Do you feel you have a good future with this Company? Yes ☐ No ☐
29. Would you be interested in selling with Victor's post war Sales Division? Yes ☐ No ☐
30. What is your reaction to being asked down to Personnel for occasional interviews? Like it ☐ Don't like it ☐
31. Do you think Victor has done a good job for the country during the war? Yes ☐ No ☐
32. Are you getting the kind of information about the Company that you want? Yes ☐ No ☐
33. What do you think of working conditions here as compared with other plants? Above average ☐ Average ☐ Below average ☐
34. How do you think your average weekly earnings (gross earnings before deductions) compare with that paid in other companys for the same type of work? Better here ☐ About the same ☐ Lower here ☐

Fig. 11.1. Part of a questionnaire used in an employee survey conducted by the Victor Adding Machine Co. (Courtesy Victor Adding Machine Co., Chicago, Ill.)

Youngberger *et al.*,[8] however, have proposed that provision also be made to obtain an indication of the *importance* of each such aspect. In particular, they suggest that the employees be asked to check each item in one of three boxes: Satisfied, Neutral, or Dissatisfied, and also check each item as being of Great Importance, Some Importance, or Little Importance.

[8] C. F. X. Youngberger, R. Hedburg, and B. Baxter, "Management Action Recommendations Based on One versus Two Dimensions of a Job Satisfaction Questionnaire," *Personnel Psychology,* 15 (1962), 145–150.

While a rank order correlation of .71 was found between results with the two scoring methods, a comparison of some of the top items showed marked differences. Table 11.3 shows a comparison of results obtained with the two methods of analysis.

Table 11.3

COMPARISON OF THE TOP SIX PROBLEM AREAS BASED ON "LIMITED" VS. "MORE COMPLETE" DATA ON JOB DISSATISFACTION

Rank	*"Limited" dissatisfaction data only*	*Rank*	*"More complete" dissatisfaction plus importance*
1	Compensation plan under which you work	1	Attitude of public toward insurance men
2	Sales programs and procedures	2	Quality of clerical assistance
3	Benefit plans in general	3	Working evenings
4	Clients relations items	4	Quality of supervision
5	Direct mail procedures	5	Services provided by home office
6	Attitude of public toward insurance men	6	Amount of paper work

COMBINATION QUESTIONNAIRES

Most typically, questionnaires are used for a particular purpose, such as the measurement of attitudes or opinions. It is possible, however, to develop questionnaires that can serve both the purposes of obtaining opinions of employees and of measuring their attitudes (or, when considered in groups, their morale). One method for doing this is described by Harris.[9] Harris identified 48 questions that would seem to indicate over-all morale from questions appearing on the Victor Adding Machine Company questionnaire illustrated in part in Fig. 11.1. Ten judges (advanced graduate students and staff members in industrial psychology) were asked to check the response to each question that "most strongly represents a favorable attitude toward the company." Any question on which the independent checks of the 10 judges did not show at least 80 per cent agreement was discarded at this point. Forty-six questions met this criterion and it is interesting to note that there was unanimous agreemen on 42 questions.

The questionnaires returned by the empoyees were then divided into two random halves after the halves had been stratified on the basis of

[9] F. J. Harris, "The Quantification of an Industrial Employee Survey, I. Method," *Journal of Applied Psychology*, 33 (1949), 103–111, also "The Quantification of an Industrial Employee Survey, II. Application," *Journal of Applied Psychology*, 33 (1949), 112–113.

several characteristics such as sex, marital status, and length of service. Item analysis against a criterion of total score was then conducted with one of the stratified halves. This resulted in the identification of 36 questions, which constituted the over-all morale scale. The reliability of this scale determined by correlating odd *vs.* even scores on the half of the papers not used in the item analysis was found to be .84.

Some typical questions making up the final scale, with the high-morale answer checked, are:

What does your family think of this Company?

Good place to work ____x____
No opinion ________
Poor place to work ________

How do you like your present job?

Very much ____x____
Pretty good ________
Not so good ________
Don't like it ________

An application of the general approach developed by Harris[10] is reported by Misa,[11] with an additional modification. He used questions, such as those illustrated below, that provide for selecting a specific response. In addition, however, he provided for optional "Remarks" for each question. The questions were grouped into three blocks, these dealing with the employee's opinions about his immediate supervisor, toward the company as a whole, and toward the income aspects of his job. For each question listed below the "high morale" response, as derived by the procedures developed by Harris,[12] is indicated by an X. The questions in each of the blocks were as follows:

Questions dealing with attitude toward the "immediate supervisor":

1. Does your supervisor (your immediate boss) seem to be interested in you and your problems?
1. (x) Yes
2. (   ) No
3. (   ) Can't say

2. Does your supervisor understand your job and the jobs he supervises?
1. (x) Yes
2. (   ) Don't know
3. (   ) No

---

10 Harris, *op. cit.*

11 K. F. Misa, *Relationship Between Three Blocks of Attitude Questions,* M. S. Thesis, Purdue University, 1963.

12 Harris, *op. cit.*

3. Does your supervisor tell you and others when you do a good job?
- *1. (x) Most of the time*
- *2. ( ) Sometimes*
- *3. ( ) Seldom or never*

4. Does your supervisor seem to get along well with the people who work for him in his department?
- *1. (x) Yes*
- *2. ( ) No*
- *3. ( ) Can't say*

5. Does your supervisor make it easy for you to express complaints?
- *1. (x) Yes*
- *2. ( ) Sometimes*
- *3. ( ) No*

6. Is your supervisor prompt in taking care of your complaints?
- *1. (x) Yes*
- *2. ( ) No*
- *3. ( ) Can't say*

7. How do you feel toward your immediate supervisor?
- *1. (x) Friendly*
- *2. ( ) No feeling*
- *3. ( ) Unfriendly*
- *4. ( ) Don't know*

Questions dealing with attitude toward the company as a whole:

8. How do you compare (company name) with other places to work in this area or city?
- *1. (x) Very good*
- *2. ( ) Average*
- *3. ( ) Not so good*
- *4. ( ) Don't know*

9. How much interest do you feel (company name) takes in you and other employees?
- *1. (x) A real interest*
- *2. ( ) Average interest*
- *3. ( ) Not much interest*
- *4. ( ) No opinion*

10. Have you ever recommended (company name) as a place to work to a friend?
- *1. (x) Yes*
- *2. ( ) No*

11. Do you think that your prospects for steady work at (company name) are?
- *1. (x) Very good*
- *2. ( ) Average*
- *3. ( ) Not so good*
- *4. ( ) Don't know*

12. Does (company name) do a good job of keeping you informed about the Company's operations?
- *1. (x) Yes, a good job*
- *2. ( ) A fair job*
- *3. ( ) Don't tell much*
- *4. ( ) No opinion*

13. Do you feel that you are informed promptly and clearly about top level management decisions that affect you?
1. (x) Yes
2. (   ) Sometimes
3. (   ) No

14. How do you feel about your opportunity for advancement with (company name) as compared with other companies?
1. (x) Better than average
2. (   ) Average
3. (   ) Poorer than average
4. (   ) Don't know

Questions dealing with attitudes toward income aspects of the job:

15. From what you know of other companies that have jobs like yours, do you feel that your take-home pay is?
1. (x) More
2. (   ) About the same
3. (   ) Less
4. (   ) Don't know

16. Do you make about the same money each month as other employees doing the same kind of work in your department?
1. (x) Yes
2. (   ) No
3. (   ) Don't know

17. Do you feel that the salary for your job, as compared with similar salaried jobs in other departments at (company name) is?
1. (x) About right
2. (   ) Low
3. (   ) Don't know

18. Considering the difference between your job and hourly paid jobs in this plant, do you feel that your take home pay is?
1. (x) About right
2. (   ) Too low
3. (   ) Can't compare
4. (   ) Don't know

19. How do you feel about (company name) salary policies as compared with other companies?
1. (x) Very good
2. (   ) Good
3. (   ) Average
4. (   ) Below average
5. (   ) Don't know

The first two blocks of questions shown above (the "Immediate Supervisor" block and the "Company" block) were developed for use with all employees in any plant. The third block of questions (the "Income" block) was developed specially for use with salaried employees in the particular plant where the survey was being conducted.

In addition to the 19 questions making up the three blocks, ten additional multiple choice questions were printed on the ballot. These additional questions dealt with other matters about which the manage-

ment wanted direct information from the salaried employees. At the end of the questionnaire was a question which read "If you have some problem or remarks to pass on to management, write it on these lines."

A general information section preceded the actual questions on the ballot. This section made it possible to break down the results by age, sex, marital status, length of service, salary, supervisor vs. non-supervisor, and department.

All salaried employees of the company were asked to assemble in certain rooms at specified times. Assurance was given that the ballots were anonymous and that the management of the company would never know which ballots were completed by which individuals. There were 769 ballots filled out and cast, representing all salaried employees who were not absent during the two days that the survey was conducted. The 769 employees casting ballots wrote in a total of 2,349 unfavorable comments on their ballots.

The analysis of the results of this survey was made in several ways. First, the over-all analysis showed what per cent of all the employees checked each response to each objective question. Examples of this analysis follow:

How do you feel toward your immediate supervisor?

Friendly	82%	checked this response
No feeling	13%	" " "
Unfriendly	3%	" " "
Don't know	1%	" " "
Left blank	1%	left this question blank

Considering the difference between your job and hourly paid jobs in the plant, do you feel that your take-home pay is:

About right	12%	checked this response
Too low	74%	" " "
Can't compare	8%	" " "
Don't know	5%	" " "
Left blank	1%	left this question blank

A second analysis involved comparing each of the 15 departments in terms of the average departmental results on each of the three blocks of questions. These blocks, as previously described, dealt with opinions about immediate supervision, the company as a whole, and income. Using the procedure described by Harris,[13] however (that takes into account the frequency of "high morale" responses), it was possible to derive an average "morale" score for each of the three blocks of questions. In addition to these three departmental measures, a fourth measure was derived from the written comments. Previous work with such written comments has shown that they can be readily and reliably classified as favorable, unfavorable, and neutral.[14]

---

[13] Harris, *op. cit.*

[14] T. M. Allen, *Analysis of Morale Survey Write-In Responses*, Ph.D. Thesis, Purdue University, 1954.

In the study being described each of the comments was so classified, and an Index of Morale from the unfavorable comments was then computed for each department. The Index used was the number of unfavorable comments per ballot cast. Toward the extremes of this Index was one department with 50 employees who wrote in only 47 unfavorable comments. The Index for this department was thus, 47/50 or .94. In another department consisting of 72 employees, 297 unfavorable comments were written in. The Index of this department was therefore 297/72 or 4.12.

There were, therefore, four departmental measures of morale for each department, i.e.

1. The average score on the immediate supervisor block of questions.
2. The average score on the company block of questions.
3. The average score on the income block of questions.
4. The average index from the written in comments.

Previous work with these four measures has shown that their reliabilities are quite acceptable for inter-department comparisons. The four sets of average scores for the 15 departments are shown in Table 11.4.

Table 11.4

DEPARTMENTAL MEANS FOR THREE BLOCKS OF QUESTIONS AND WRITTEN COMMENTS *

Dept. No.	*Supervisor* **	*Company* **	*Income* **	*Written-in comments****
1	4.98(4)	3.38(2)	1.18(5)	.94(2)
2	3.98(14)	2.12(13)	.80(12)	4.40(13.5)
3	5.23(3)	2.72(7)	1.76(2)	2.15(7)
4	5.37(2)	3.93(1)	1.42(4)	1.17(3)
5	4.56(9)	2.64(8)	1.60(3)	2.00(5)
6	4.79(6)	3.22(3)	1.13(6)	2.34(8)
7	3.99(13)	2.25(10)	.69(15)	4.12(12)
8	4.68(7)	2.38(9)	.82(13)	3.70(9)
9	3.73(15)	2.15(12)	1.03(8.5)	5.23(15)
10	4.46(10)	2.74(6)	1.03(8.5)	3.75(10)
11	6.22(1)	3.11(4)	2.17(1)	.67(1)
12	4.72(8)	2.79(5)	1.08(7)	1.67(4)
13	3.82(5)	1.78(15)	.72(14)	4.29(13.5)
14	4.22(12)	1.88(14)	.81(11)	3.79(11)
15	4.43(11)	2.23(11)	1.00(10)	2.03(6)

* *Rank order shown in parentheses.*
** *High morale is shown by a large number.*
*** *High morale is shown by a low number.*

Table 11.4 quickly shows those departments where morale is high, average, or low in each of the respects tapped by the four departmental measures. For example, Department No. 11 is best in three of the meas-

ures and fourth from the top in the fourth measure (questions about the company). Departments 2 and 14 are far below average in all four measures.

When the low departments have been identified from Table 11.4, it is enlightening to read the comments that were written by employees in these departments. For example, here are a few of the 220 unfavorable comments written by the 50 employees in Dept. No. 2.

Steady work but no advances
Lack of coordination in the organization
Some supervisors do not really know the people under them
I see little opportunity for advancement in my department
Our supervisor doesn't care about anything

(There were 215 more unfavorable comments written by the 50 employees in this department.)

With specific information of this sort, management can go into action to train or replace supervisors and either change company policy or make clear to the employees why the present policies have been established.

The intercorrelations of the four measures by departments are also of interest. The Rank-order Correlations are shown in Table 11.5.

Table 11.5

RANK-ORDER CORRELATIONS BETWEEN DEPARTMENT AVERAGE SCORES ON THE FOUR ATTITUDE MEASURES SUMMARIZED IN TABLE 11.4

	Company block	*Income block*	*Written comments*
Supervisor Block	.69	.62	.71
Company Block		.71	.80
Income Block			.75

It will be noted in Table 11.5 that the written in comments quite effectively tap the attitude characteristics that are measured by the three blocks of questions. In addition, of course, the written comments specifically mention the things that the employees do not like or that are sources of dissatisfaction to them.

In addition to the inter-department comparisons from statistical analysis, the final report to management included all of the written in comments arranged by the departments from which they came. These quotations constituted the body of the report, requiring 189 typed pages. Management found these comments very informative and considered them just as meaningful as the more detailed and quantified statistical results.

Finally, an over-all summary of the most freqnent unfavorable comments was found very helpful by management. These comments are summarized for all salaried employees in Table 11.6.

Table 11.6

SUMMARY OF TOPICS MENTIONED UNFAVORABLY TEN OR MORE TIMES

	No. of mentions
Cafeteria food poor and/or expensive	46
Merit increases too slow or inconsistent	13
Hourly wages out of line with salaries	85
Too much hiring from outside	25
Work space crowded, noisy, hot, etc.	15
Poor cooperation between departments or divisions	13
Classification should be reviewed	22
Poor communication of policy and plans	72
Organization of department poor	36
Supervision poor	61
Coffee break needed	11
Better equipment needed	12
Sick leave and vacation should be cumulative	13
Foremen should have more authority	10
Management should take a firmer stand with union	10

"THEME" ANALYSIS OF EMPLOYEE COMMUNICATIONS

A very promising method of measuring the reaction of employees to various aspects of their jobs has been developed by the Employee Research Section of the General Motors Corporation.[15] This work made use of 174,854 letters entered in the General Motors MJC (My Job and Why I Like It Contest). Letters were entered by 58.8 per cent of all eligible employees. Aside from the evaluation of letters for awarding prizes, the letters proved to be a source of valuable information regarding opinions of employees in the various divisions. To "mine" this source of information, a *theme analysis* of the letters was made. Some of the letters were reviewed for content to identify the "themes" that were being mentioned. Fifty-eight themes were identified. Some examples are: supervision, wages, pride in product, security, employee relations, and pension plans.

After identification of the themes, each letter was analyzed in order to determine what theme or themes it mentioned. An example will illustrate how this was done. Below are excerpts from one letter, with the corresponding theme for each of three portions of the letter.

[15] C. E. Evans and L. N. Laseau, "My Job Contest," *Personnel Psychology Monograph No. 1* (Washington, D.C., 1950).

EXCERPTS FROM LETTER	THEME
My job with General Motors has given me what every man wants—*security for his family*	*Security theme*
In less than a month I was getting top rate	*Success theme*
Knowing that a *chance for advancement* is possible	*Opportunity for advancement theme*

Although the participants would be expected to mention only the *positive* factors of their jobs, it was believed that their *lack of mention* of certain factors might be meaningful. Among the various analyses that were made, one consisted of a rank order of the frequency of mention of themes in the letters from the various divisions. A comparison was then made of the rank order of frequency of themes mentioned in *each* division with the rank order for *all* divisions. An illustration of this comparison is given in Table 11.7. This table shows the 20 themes that were mentioned most frequently for all divisions, giving their rank order of frequency of mention. The table also shows the rank order of those same 20 themes from one particular division (No. 48). Although the first six themes as mentioned by employees in Division 48 had the same rank order of frequency of mention as in the corporation as a whole, it can be

Table 11.7

RANK ORDER OF SELECTED THEMES FROM GENERAL MOTORS "MY JOB CONTEST" FOR ALL DIVISIONS AND FOR DIVISION 48

	Rank order	
Theme name	*All divisions*	*Division 48*
Supervision	1	1
Associates	2	2
Wages	3	3
Work type	4	4
Pride in company	5	5
Management	6	6
Insurance	7	9
Pride in product	8	11
Benefits from wages	9	13
Pride in stability of company	10	12
Safety	11	16
Security	12	10
Training, education, experience	13	7
Opportunity for advancement	14	8
Medical facilities	15	23
Teamwork	16	14
Tools, methods, equipment	17	17
Paid vacation	18	20
Cleanliness	19	24
Pride in important work	20	15

Source: *Adapted from Evans and Laseau,* op. cit.

seen that the rank orders of some of the subsequent themes differed, some of them rather markedly. It can be seen, for example, that *opportunity for advancement* was ranked fourteenth by all divisions collectively, but was ranked eighth by employees in Division 48; this suggests that the employees in that division perceive their opportunities for advancement to be considerably better than do employees in the company at large. On the other hand, the *medical facilities* in that division are perceived as being below average, as indicated by the fact that for all divisions this theme was fifteenth whereas for Division 48 it ranked twenty-third.

This General Motors approach to theme analysis of employee communications presents some very promising possibilities for the measurement of employee reaction toward various aspects of their jobs.

INTERVIEWS

Still another method of obtaining information about personnel reactions is the use of the interview.

Interviews with Employees In some cases interviews are carried out with present employees, either all employees or a sample. The employees are given assurance that the information furnished will not be used for any administrative purpose. To provide such assurance, the interviews usually are conducted by representatives of some outside organization, such as a consulting firm or a university.

Such interviews may be *guided* or *unguided*. In the guided interviews, the interviewer asks a series of questions, each of which may be answered by a simple "yes," or "no," or by some other word or phrase. In the unguided interview, the interviewer asks more general questions to encourage the employee to express himself in whatever manner he wishes. The comments of the employee are used by the interviewer as the basis for assessing the employee's attitudes and opinions.

The guided interview is something like an orally administered questionnaire, the results of which usually can be summarized statistically. The results of the unguided interview, on the other hand, do not lend themselves as readily to quantification. With this method the interviewer usually "rates" the attitudes of employees. In this connection there is some evidence to indicate [16] that such ratings tend to overestimate or underestimate the strength of employee attitudes, and are highly inconsistent.

Aside from such limitations of the interview, it must also be recognized that the cost of this method is high. Except under unusual circum-

---

[16] "Comparison of Scaled Questionnaire and Interview Methods in Appraising Industrial Attitudes," *Research Bulletin 1*, Bureau of Industrial Psychology, University Extension Division, University of Wisconsin (undated).

stances, therefore, a questionnaire usually provides a more adequate and practical method of measuring employee attitudes and opinions.

Exit Interview The exit interview also provides the opportunity to obtain some inklings about employee attitudes toward the company, and feelings or opinions with regard to specific practices, policies, working conditions, etc. Aside from its value as a barometer of employee attitudes and opinions, it has other values, of course. For example, the check it provides on policies covering employment, placement, training, grievances, and wages helps to identify sources of employee dissatisfaction and reasons for high turnover.

An employee who is working for the company often is reluctant in expressing his true feelings. Also, many managements feel that it would not be good industrial relations practice to encourage present employees to give verbal expression to their feelings. But the employee who has terminated his employment usually is quite willing to say what he thinks about the company, and management has little to lose (and often much to gain) by listening attentively to what he has to say. This fact has resulted in an increasing use of the exit interview—an interview conducted with the employee leaving the company. To obtain the greatest value from the exit interview the personnel man should conduct the interview in private and should take all the time needed to allow the terminating employee to tell his full story in his own words. Any single employee may have grievances for which there is no real basis in plant practices, yet if an appreciable number of employees terminating their employment mention the same situations or practices as unsatisfactory, it is usually safe to conclude that there is a real reason for their discontent. Under such circumstances, it is also a reasonable assumption that employees still on the job are not entirely satisfied with these practices or policies. Information obtained in a series of exit interviews should not, of course, be automatically followed by changes in managerial policy whenever a source of grievance is revealed, but most managements now believe that such information should at least be considered when making future managerial decisions.

12

Motivation and Job Satisfaction

One of the most enigmatic aspects of the management of industrial and business organizations is that relating to human motivation. The manifestations of the apparent vagaries of human motivation are reflected, on the one hand, by such circumstances as low production, wildcat strikes, personal conflicts between supervisors and subordinates, restriction of output, absenteeism, and high turnover, and, on the other hand, by such circumstances as individual creativity, outstanding organizational achievements, high *esprit de corps,* and the strong personal commitment of individuals to their organization that generates work effort above and beyond the call of duty.

CAUSATION OF HUMAN BEHAVIOR

In view of such varied behavior, a basic assumption must be made, namely, that all behavior is caused—that there are certain causal influences that bring about the behavior. In the jargon of the theoretical and experimental psychologists, a *stimulus* acts upon the *organism* to produce *behavior* (a response). A simple illustration of this is the instance of a person touching a hot stove. The stimulus (the hot stove) acts upon

the organism (the person) to produce behavior (to remove the hand). Most of the situations in everyday life, however, are much more complex than this. For example, when driving in traffic the driver has many "stimuli" in his environment, such as other cars, pedestrians, the traffic lights, the road and road conditions, and perhaps a back-seat driver. His driving behavior is the consequence of an *interaction* of the stimuli and the driver. The driver himself has had previous driving experiences which will influence his behavior in a given situation. Further, his *perception* of the stimuli may influence his behavior. For example, if the driver perceives the speed of a car coming toward him as being less than it actually is, he will behave as though it were travelling at the slower speed; for example, he may then attempt to pass a car in front of him.

In a somewhat parallel vein, a clerk in the complaint department of a store may look at the long line of customers, realize that the closing time is coming (before which she cannot possibly handle all the customers), and may "perceive" their complaints to be directed toward her personally rather than toward the store. Her resulting behavior may be to act in a very snippy and discourteous way to the customers, which, of course, may cause them to take out their wrath directly on her, thereby aggravating the whole situation.

In this process of responding to stimuli, it is very easy to see how the previous experiences of the individual and the attitudes of the individual can affect the resulting behavior. Previous experiences, especially those which have been repeated often enough to form behavior habits, can have the effect of causing a person to persist in a particular type of behavior in situations of a given type. Attitudes, in turn, can affect our perceptions of events or objects in our environment and can provide a predisposition to react in a particular manner.

HUMAN MOTIVATION

Central to the way in which an individual behaves in any given circumstance is the nature of his motivation. In order better to understand individual motivation, it is helpful to have some understanding of some of the more general facts and principles relating to motivation. In this connection let us first differentiate between *needs* and *incentives*. *Needs* are the internal, felt wants of individuals. They are also referred to as *drives* or *desires*. *Incentives* (also referred to as goals) are external factors which the individual perceives (rightly or wrongly) as possible satisfiers of his felt needs. Thus, *thirst* is a *need*, and *water* is an *incentive*. Water (the incentive) satisfies thirst (the need).

Basic Needs It has been suggested by Maslow[1] that sound motiva-

[1] A. H. Maslow, *Motivation and Personality* (New York: Harper & Row, Publishers, 1954), p. 69.

tional theory should assume that people typically are continuously in a motivational state, but the nature of the motivation is fluctuating and complex; further, human beings rarely reach a state of complete satisfaction except for a short time. As one desire becomes satisfied, another arises to take its place, and as this desire becomes satisfied, another replaces it. This never-ending sequence gives rise to Maslow's theory of motivation in which a hierarchy of needs is postulated. This theory has as its central feature the concept of human needs being generally ordered in terms of their relative potency as human motivators. Some modifications of Maslow's original categories (especially in terms of labels) have been made by McGregor.[2] These are described and discussed briefly.

The physiological needs. The physiological needs are taken as the starting point, and are conceived to be the most prepotent of all. To the person in a state of virtual starvation or water deprivation, matters other than food or water are of little concern. He could not care less. In some countries of the world, food deprivation is indeed a dominant motivator. In other countries, however, there is little actual starvation, and for most people the basic physiological needs are fulfilled. Even when we say we are hungry, we are not really seriously deprived of food. Frequently this "need" becomes intertwined with other nonphysiological needs, such as those associated with comfort or dependence. The expressed "need" for a sandwich or cup of coffee may really be based more on a desire to avoid continuing with one's job, or a desire to have a chat with a buddy.

The safety needs. Once the physiological needs are relatively well met, there emerges a new set of needs which are categorized generally as safety needs. These generally are concerned with protection against danger, threat, and deprivation. Protection against physical dangers are of less consequence now (in our civilization) than was the case in the past. On the other hand, in an industrial society the safety needs may take on considerable importance in the context of the dependent relationship of employees to employers. As pointed out by McGregor,[3] the safety needs may serve as motivators in such circumstances as arbitrary management actions, behavior which arouses uncertainty with respect to continued employment, and unpredictable administration of policy.

Social needs. Once the physiological and safety needs are reasonably well fulfilled, the social needs become important motivators of behavior—needs for belonging, for association, for love, for acceptance by his fellows, and for giving and receiving friendship.

The ego needs. Next in the hierarchy are the ego needs. McGregor[4]

---

[2] D. M. McGregor, "The Human Side of Enterprise," *Management Review,* 46 (No. 11, November, 1957), 22–29, 88–92.

[3] McGregor, *op. cit.*

[4] McGregor, *op. cit.*

distinguishes two kinds, as follows: (1) those needs that relate to one's self-esteem—needs for self-confidence, for achievement, for competence, for knowledge; and (2) those needs that relate to one's reputation—needs for status, for recognition, for appreciation, for the deserved respect of one's fellows. In contrast with the lower needs, the ego needs are rarely fully satisfied. These needs, however, usually do not become dominant until the lower needs have been reasonably fulfilled.

Self-fulfillment needs. Highest on the postulated totem pole of needs is that of self-fulfillment—of realizing one's own potentialities, for continual self-development.

Discussion Keeping constantly in mind that the above hierarchy represents the *general* order of relative potency of the various needs (and not that of all *individuals*) Maslow[5] points out that the hierarchy is characterized by some supporting aspects or features, a few of which are given below:

1. The higher needs are a later evolutionary development.
2. The higher the need, and the less imperative it is for sheer survival, the longer gratification can be postponed, and the easier it is for the need to disappear permanently.
3. Living at the higher need level means greater biological efficiency, greater longevity, less disease, better sleep, appetite, etc.
4. Higher needs are less urgent, subjectively.
5. Higher need gratifications produce more desirable subjective results, i.e., more profound happiness, serenity, and richness of the inner life.
6. Pursuit and gratification of higher needs represent a general healthward trend.
7. Higher needs require better outside conditions (economic, education, etc.) to make them possible.
8. Satisfaction of higher needs is closer to self-actualization than is lower-need satisfaction.

Maslow suggests that the various levels are interdependent and overlapping, each higher-level need emerging before the lower-level need has been completely satisfied. In addition, he points out that individuals may jumble the "order" and importance around.

It should be noted here that while there is reasonable support for the hypothesis that to some extent human needs do have some hierarchical order, questions have been raised regarding the generality of Maslow's formulation. Herzberg, Mausner, and Snyderman,[6] for example, seriously question the theory that motivation in human work is primarily predicated on a hierarchy of needs.

---

[5] Maslow, *op. cit.*, Chap. 8.

[6] F. Herzberg, B. Mausner, and B. B. Snyderman, *The Motivation to Work* (New York: John Wiley & Sons, Inc., 1959).

Incentives The needs of individuals, then, serve as driving forces in human behavior. As indicated earlier, the specific objects or behaviors that are perceived as being able to satisfy these needs are *incentives.* Incentives can be either positive or negative. A positive incentive is one that attracts a person like a magnet, as, for example, the prospect of a promotion. On the other hand, there are some consequences or events that people seek to avoid; these are negative incentives. Thus, an employee continues to work at a job that he does not like because of the loss of pay he would experience if he did not go to work.

One other point should be made about incentives, this being the distinction between *real* and *substitute* incentives. Given a particular need, there typically is a certain incentive that most logically would fulfill that need. Frequently, however, it is not possible to achieve that particular incentive, or its costs (in terms of other satisfactions) might be excessive. Thus, a person may visualize some *other* incentive as being a substitute for the "real" one. A man may fail to achieve a particular goal, such as becoming a doctor, and redirect his energies toward becoming a medical technician. Or an employee may have his heart set on becoming a foreman, but when this goal fails to materialize, he may become more active in his lodge and ultimately end up as its president, thus fulfilling his need for status.

REACTION TO FRUSTRATION

Frustration occurs when some obstacle is placed in the direct path of an individual's movement toward a goal (incentive). As defined by Costello and Zalkind,[7] frustration refers to the event rather than to the internal feeling state. Frustrating events are, of course, part and parcel of every phase of life—in family life, in schools, in social situations, in jobs. In whatever circumstance, frustrating events give rise to some type of behavior. The behavior that is induced by frustration varies with the individual and the circumstance, and can range over a wide gamut.

Positive Reactions Frustration in some circumstances leads to positive, constructive resolution of the circumstances that block the achievement of the goal. This effect presumably occurs by reason of increased drive that is then directed toward problem solving, in particular, trying to get around the barrier that prevented the achievement of the original goal. It is probable that many of the scientific, technological, and cultural developments of history, as well as many less spectacular developments, have been stimulated by frustrating events in the lives of the individuals. In other situations the frustration may cause the individual

[7] T. W. Costello and S. S. Zalkind, *Psychology in Administration* (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1963), p. 131.

to divert his energies into an entirely different type of activity, but one that is also constructive in its nature.

Negative Reactions On the other hand, frustration can generate various forms of nonconstructive behavior such as aggression, regression, fixation, resignation, negativism, repression, and withdrawal, as well as others.[8]

Aggression. This is a form of behavior that is characterized by some type of attack toward another person or object. The four-year-old boy who beats his playmate over the head with a baseball bat is obviously exhibiting aggressive behavior. The forms of aggression of adults can be very overt, physical or verbal. They can also be very disguised; the titbit of shop gossip, or the subtle comment by an aspirant for a promotion about one of his possible competitors, can be just as devastating to the victim as the baseball bat.

Regression. Regression is the tendency for an individual in a frustrating circumstance to revert to an earlier form of behavior, such as putting on a temper tantrum or pouting. Such behavior most likely would occur in the case of an individual who, during his formative years, found that such behavior worked; in other words, he got his way by, say, a display of temper.

Fixation. In fixated behavior, the individual persists in a particular kind of nonconstructive behavior even though it is virtually manifest that the behavior is inadequate to resolve the problem. Thus, a mechanic may persist in trying to fit a bolt in place, even though he realizes that it is the wrong size.

Resignation. Resignation is the tendency for an individual to give up—to withhold any sense of emotional or personal involvement in the situation.[9] Failing to achieve some goal, the person loses any positive concern about his job and adopts an apathetic attitude toward the situation.

Negativism. In this form of behavior, the individual adopts a negative, resistive frame of reference with respect to the situation. A person whose suggestions have not been accepted, for example, may take a dim view of any other ideas.

Repression. This behavior is characterized by blocking out from consciousness those cognitive associations that are disturbing. It is an unrealistic form of behavior since it implies that the problem will simply go away if one doesn't think about it.

Withdrawal. In this type of behavior, the frustrated individual simply removes himself from the situation in question—either physically or psychologically. A person who is not able to cope with a business adversary may avoid situations which would put him in contact with

---

8 *Ibid.*, pp. 148–149.
9 *Ibid.*, p. 149.

the individual. An individual who is the butt of jibes and jokes may become a "loner."

Discussion There are numerous other reactions to frustration, such as fantasy and rationalization. In whatever form, reactions to frustration are efforts on the part of the individual to resolve the circumstance. There are at least a couple of implications that might be mentioned. First, the way in which an individual behaves has implications in terms of his own personal adjustment. In some accentuated cases, the maladjustive behavior may be indicative of serious emotional disturbance that requires professional attention. In the second place, the behavior of individuals with respect to frustration has implications relating to their job performance. Supervisors, executives, or others can perhaps deal more effectively and constructively with individuals whose job performance under frustration is undesirable if they are able to understand the underlying basis of the behavior.

REACTION TO INTERNAL CONFLICT

What is called frustration stems from events in which individuals are prevented from achieving their goals. It is, then, the consequence of a conflict between the individual and his environment—the world about him. On the other hand, there can also be conflicts internal to the individual. These, too, enter into the motivation of people, and therefore have implications relating to human behavior.

Cognitive Dissonance In recent years the concept of *cognitive dissonance* has been applied to such internal conflicts. This concept, crystallized by Festinger,[10] deals with the degree of "consistency" that characterizes an individual's opinions, attitudes, and behavior. In general, an individual strives to maintain reasonable consistency in these matters, since a high degree of inconsistency usually causes a person to feel "uncomfortable."

In his theory he uses the terms *consonance* (instead of *consistency*) and *dissonance* (instead of *inconsistency*). These terms generally refer to the degree of congruence, and non-congruence, between two aspects of one's psychological environment; these aspects can be attitudes, beliefs, perceptions, opinions, or behaviors in one's situation. Dissonance exists where—for some reason or another—there is a significant disparity between elements of a situation which are in some respect incompatible. For example, if a supervisor believes basically in treating his subordinates courteously, but is expected by his management to rule with an iron hand —and does so—his behavior (being "forced" upon him by his superiors)

[10] L. Festinger, *A Theory of Cognitive Dissonance* (New York: Harper and Row, Publishers, 1957).

is at odds with his own personal beliefs and convictions. This disparity between his beliefs and his (forced) behavior would be the source of his dissonance.

Variables that influence dissonance. Whether dissonance will occur in a given situation, and its extent, depends upon a number of variables. One of these is the *importance* of the relevant aspects of the situation, as perceived by the individual. Another factor is the *extent of choice* that is available. Where there is only one course of action that is available to an individual (even a course that is intrinsically distasteful to him), dissonance usually would be less than if two or more equally distasteful courses are available to him. In addition, the *magnitude of the reward* —the potential gains to the individual—enters into the equation. Where, for example, the rewards of two possible courses of action are both high, the dissonance that might occur typically would be greater after a choice is made, than where the gains from the two courses of action are relatively nominal.

Situations in which dissonance occurs. Some of the events and circumstances of everyday life inevitably introduce disparities of one sort or another.[11] While there are various types of circumstances that are dissonance-producing, only two will be mentioned here. One type relates to disparities between one's personal, private beliefs and attitudes, as contrasted to those which the individual for one reason or another presents to the public. An employee, for example, may personally approve of some policy of management, but "publicly" indicate disapproval because this is the prevailing attitude of his fellow workers.

Another type of circumstance in which dissonance may occur is exemplified when a choice must be made from among two or more alternatives that are about equally desirable (or undesirable). A decision to choose one may later raise questions in one's mind as to whether the choice that was made was really the best one. For example, if an individual has accepted one position instead of another, he may start to add up the pros and cons of the one he did choose as compared to the one he rejected.

Dissonance may also be created as the consequence of new information. Thus, a sales manager's opinion about the potential market for a new product might be at odds with the results of a market-research survey that indicates that the product would not have wide acceptance.

Reaction to dissonance. When an individual experiences dissonance, he tends to reduce it in one way or another. The recognition of the disparity may lead to any one of several "solutions" that tend to minimize its magnitude. Depending on the type of circumstance in question, some methods of resolution are: enhancing a chosen action

---

11 Festinger, *ibid.*

by perceiving it in a more favorable light; depreciating a rejected choice; magnifying the rewards that are associated with a chosen action; adopting an attitude of apathy; tolerating the disparity; seeking additional information to bring to bear on the matter; and by questioning the "facts" that are at odds with one's opinions or behavior. This last method of reducing the dissonance is illustrated by the results of a survey of opinions about the linkage between cigarette smoking and lung cancer. The results of this survey are given in Table 12.1. This shows that, in general, more heavy and moderate smokers were of the opinion that the linkage had not been proved, than were nonsmokers or light smokers.

Discussion It has been argued by Festinger [12] that cognitive dissonance is an important determinant of human behavior. In particular, he hypothesizes that the existence of dissonance is comparable to any other need state in that it gives rise to behavior that is directed toward eliminating it. Successful reduction is rewarding in the same sense that eating when one is hungry is rewarding.

Table 12.1

OPINIONS OF RESPONDENTS CONCERNING THE LINKAGE BETWEEN CIGARETTE SMOKING AND LUNG CANCER: AN EXAMPLE OF COGNITIVE DISSONANCE *

	Per cent who thought linkage was		
	Proved	*Not proved*	*No opinion*
Nonsmokers (n = 348)	29	55	16
Light smokers (n = 59)	20	68	12
Moderate smokers (n = 105)	16	75	9
Heavy smokers (n = 41)	7	86	7

**From* Minneapolis Sunday Tribune, *March 21, 1954. Courtesy of the* Minnesota Poll *and the* Minneapolis Sunday Tribune.

In the industrial setting, discrepancies inevitably arise, such as in labor-management relations, in policy-determination, and in personal relations between and among individuals. While such conflicts can serve as externally frustrating events, they also can generate internal dissonance within individuals. Thus, a supervisor may be obligated to place into effect a policy or procedure of which he does not approve.

An understanding of the influence of cognitive dissonance in the behavior of others is helpful in dealing with the consequences of it. There are, of course, many different ways in which dissonance can influence the behavior of individuals. In this connection, dissonance may

[12] L. Festinger, "The Motivating Effect of Cognitive Dissonance" in *Assessment of Human Motives,* G. Lindzey, ed. (New York: Holt, Rinehart & Winston, Inc., 1958), pp. 65–86.

actually force persons to look more favorably on their activities and even to be more productive in their accomplishment. Weick,[13] for example, has demonstrated that persons who strongly dislike their supervisor but agree to perform a task which he assigns, are significantly more productive than persons who comply with a more neutral task-setter. This is assumed to occur because the person justifies his unwarranted compliance by enhancing the value of the task.

Dissonance can lead either to constructive actions, or to actions that are undesirable in terms of the organization. On the positive side of the ledger, for example, it can lead to rational analysis of choices, or to the collection of more information for use in making decisions. On the other hand, it can generate apathy, it can lead to the avoidance of making any decision, it can cause people to be less critical in analyzing problems, or it can contribute to political maneuvering toward the end of gaining support for one's point of view.

While it is not realistic to hope to create situations in which dissonance would not develop, it is desirable to take whatever steps are feasible that would tend to minimize it and that would reduce its undesirable consequences. Some of the actions that are discussed in the next chapter would have the effects of creating situations that would be conducive to these ends.

ATTITUDE FORMATION

The manner in which attitudes become formed, and the process by which they are modified, are as yet not fully understood. As pointed out by Katz,[14] there have been two opposing views relating to attitude formation. One of these assumes an irrational model of man—that man's attitudes stem primarily from emotional forces and appeals to self-interest. The other assumes that man is essentially rational—that he seeks understanding, and typically exercises his reasoning and discriminating abilities. Katz[15] maintains that neither of these positions can be fully supported. In turn, he postulates that the formation of attitudes stems from the psychological needs that give rise to them, and, in particular, suggests that they serve four functions. These are described briefly below.

The Adjustment Function This function is essentially a reflection of the fact that people strive to maximize the rewards in their external

---

[13] K. E. Weick, "Reduction of Cognitive Dissonance through Task Enhancement and Effort Expenditure," *Journal of Abnormal and Social Psychology*, 68 (1964), 533–539.

[14] D. Katz, "The Functional Approach to the Study of Attitudes" and "Determinants of Attitude Arousal and Attitude Change," *Public Opinion Quarterly*, 24 (1960), 163–176, 176–192.

[15] *Ibid.*

environment and to minimize the penalties. Thus, individuals tend to adopt those attitudes that are compatible with their perceptions of the utility of the attitudinal "object," i.e., the specific objects, people, or symbols with which the attitude is associated.

The Ego-defensive Function When there is some form of threat to the ego, either from external sources or from internally induced anxieties, the individual may generate those attitudes that "protect" the ego. Thus, a sarcastic remark by a fellow worker might cause an individual to adopt a defensive or aggressive attitude toward the one making the remark.

The Value-expressive Function To some degree individuals may adopt those attitudes that are consonant with their central values and with their self-perceptions.

The Knowledge Function Still another function that is served by adoption of attitudes is the need for understanding and knowledge that form a consistent and meaningful framework. As indicated earlier in the discussion of cognitive dissonance, one method of reducing dissonance is by modifying one's attitudes in the light of new information or events.

MOTIVATION IN WORK

So far, we have been discussing human motivation in general terms. Let us now turn our attention more specifically to the question of motivation in work situations, and some of the related aspects of job satisfaction, attitudes toward work, and the relationship between such variables and work behavior. In human work, as well as in other aspects of life, the motivation of people is, of course, central to any form of behavior. Even the slaves of bygone empires were "motivated" to work by, say, threat of punishment or deprivation of food and water. In today's industrial economy even an individual who despises his job may continue at it since the alternatives are more distasteful—such as not being able to buy food, clothing, and shelter.

Satisfactions and Dissatisfactions As indicated earlier, while human motivation is characterized by basic "needs," it is specifically directed toward achieving certain desirable, positive goals or incentives, or, conversely, toward avoiding other undesirable, negative consequences. Thus, in the work setting the question arises as to what variables are perceived (within the value systems of individuals) as being desirable goals to achieve, and conversely, undesirable conditions to avoid.

In this connection, Herzberg and his associates [16] have carried out a very significant study that sheds some light on this question. This study consisted of an intensive analysis of the experiences and feelings

---

[16] F. Herzberg *et al., The Motivation to Work* (New York: John Wiley & Sons, Inc., 1959).

of two hundred engineers and accountants in nine different companies. During structured interviews they were asked to describe a few previous job experiences in which they felt "exceptionally good" or "exceptionally bad" about their jobs. They were also asked to rate the degree to which their feelings had been influenced—for better or worse—by each experience which they described.

Analysis of interview data The recorded interview data were broken down into "thought units," each of which related to a single event or condition that led to a feeling, a single characterization of a feeling, or a description of a single event. A few examples are given below: [17]

1. The way it was given to me showed that the supervisor had confidence in my work.
2. Feel fresh and eager, ready to come to work.
3. Gave me an attitude of indifference toward my job, didn't care whether it got done or not.
4. Wasted time doing unnecessary tasks. After the job, I knew it wouldn't work, just sat there until he came back for it.
5. I like to know there's a reason for doing the job.

Five thousand such statements were classified into categories such as those given in Fig. 12.1. Within each such category there were subcategories that provided for various specific kinds and degrees of responses —both positive and negative.

A major phase of the analysis consisted of various comparisons between the "high" job-attitude and "low" job-attitude "sequences." (A sequence was any one of the job experiences that were described during the interviews; these were divided into those in which "high" job-attitudes and "low" job-attitudes were expressed.)

Results of analysis. Some of the most significant results are given in Fig. 12.1. This shows, for the high and for the low job-attitude sequences, the percentage in which each of the various categories (factors) appeared —as the "thought units" were classified. The major inferences of these (and other) data from the study relate to the distinction between *satisfiers* and *dissatisfiers*. The categories that are primarily associated with high job-attitudes generally are associated directly or indirectly with the *job activities* as such, these categories being achievement, recognition, the work itself, responsibilities, and advancement. Since positive expressions relating to these factors are generally associated with high job-attitude situations, they are referred to as *satisfiers*. On the other hand, the factor categories that were associated with low job-attitude situations are those that are extrinsic to the work itself, that are primarily associated with the *job context* rather than with the job activities; the more important of these are company policy and administration, technical

[17] *Ibid.*, p. 38.

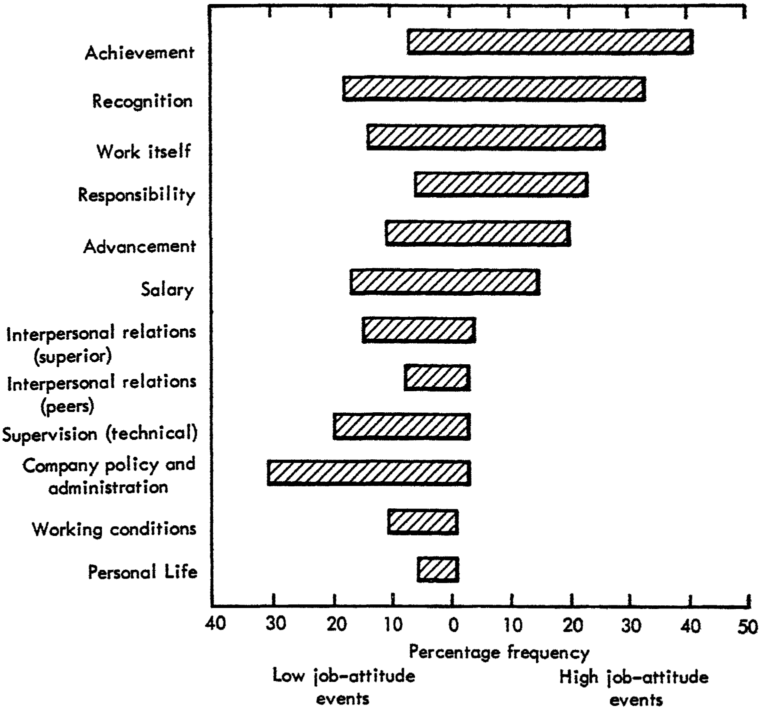


Fig. 12.1. Percentage of "high" and of "low" job-attitude sequences in which each of the categories appeared. (Adapted from Herzberg, *et al., op. cit.*, p. 72.)

supervision, interpersonal relations (supervision), and working conditions. Generally negative feelings regarding such factors dominated the reactions of people to the low job-attitude experiences they reported.

These results have been interpreted by Herzberg as indicating that the factors that lead to positive job attitudes do so primarily because they satisfy the individual's need for self-actualization in his work. The conditions that surround the job cannot give him this basic satisfaction, although (when these conditions are satisfactory) they can meet the needs of the individual to avoid unpleasant situations. Thus, one group of factors (those revolving around the job itself) serve the need to develop in one's occupation as a source of personal growth. The second group (those associated with the job conditions) operate as an essential base to the first. While they do not typically bring about positive satisfaction, they can operate to prevent dissatisfaction.

Personality Differences The potency of any of the various factors dealt with in Herzberg's analysis is not entirely a function of the factors them-

selves. It is also influenced by the personality characteristics of the individual. In this connection, Herzberg[18] refers to "motivation seekers" and "maintenance seekers." The motivation seekers generally are individuals who are primarily motivated by the "satisfiers" such as achievement, responsibility, advancement, and other factors associated with the work itself. On the other hand, the maintenance seekers tend to be more concerned with the factors surrounding the job, such as supervision, working conditions, pay, and company policy, and tend to be less concerned with the kind and quality of their work.

Related Studies Substantial confirmation of Herzberg's findings regarding satisfiers and dissatisfiers in the work situation comes from a subsequent study by Schwartz, Jenusaitis, and Stark[19] with a group of supervisors in the utility industry. Another study of motivation that was patterned after Herzberg's has also been carried out at Texas Instruments. This study, reported by Myers,[20] involved the intensive interviewing of 282 subjects (including 52 females) in the following groups: scientists, engineers, manufacturing supervisors, hourly technicians, and female assemblers. The interview data were analyzed in much the same way, and the results, in general, tended to confirm the results and theories from Herzberg's study. It should be added, however, that in the study by Myers[21] there were noticeable differences in the specific factors that were perceived as being important by those in the various groups. Some of the more important factors (both positive satisfiers and negative dissatisfiers) are given below for the occupational groups:

	Satisfiers	*Dissatisfiers*
Scientists:	Work itself	Responsibility (disappointments)
Engineers.	Advancement Work itself	Responsibility (disappointments) Company policy and administration
Supervisors:	Advancement Responsibility	Pay
Technicians:	Advancement Responsibility	Work itself Pay
Female assemblers·	Supervision	Recognition (lack) Security (lack)

While factors associated with the work activities as such (advancement, recognition, responsibility, the work itself, etc.) generally are considered as *satisfiers*, unsatisfactory conditions relating to such factors can operate to induce dissatisfaction. In the case of the technicians, for exam-

18 *Ibid.*

19 M. M. Schwartz, E. Jenusaitis, and H. Stark, "Motivational Factors among Supervisors in the Utility Industry," *Personnel Psychology*, 16 (1963), 45–53.

20 M. S. Myers, "Who Are Your Motivated Workers?" *Harvard Business Review*, 42 (Jan.–Feb., 1964), 73–88.

ple, it was suggested that the "work itself" served as a dissatifier since such workers might feel that they get the dirty work of the scientists.

The analysis reported by Myers [22] did not stop there, however. Rather, the company sought to put into practice some of the implications of the research. In particular, these efforts were directed toward satisfying both the maintenance needs and motivational needs of employees. Fig. 12.2 depicts these two types of needs in graphic form. In its attack

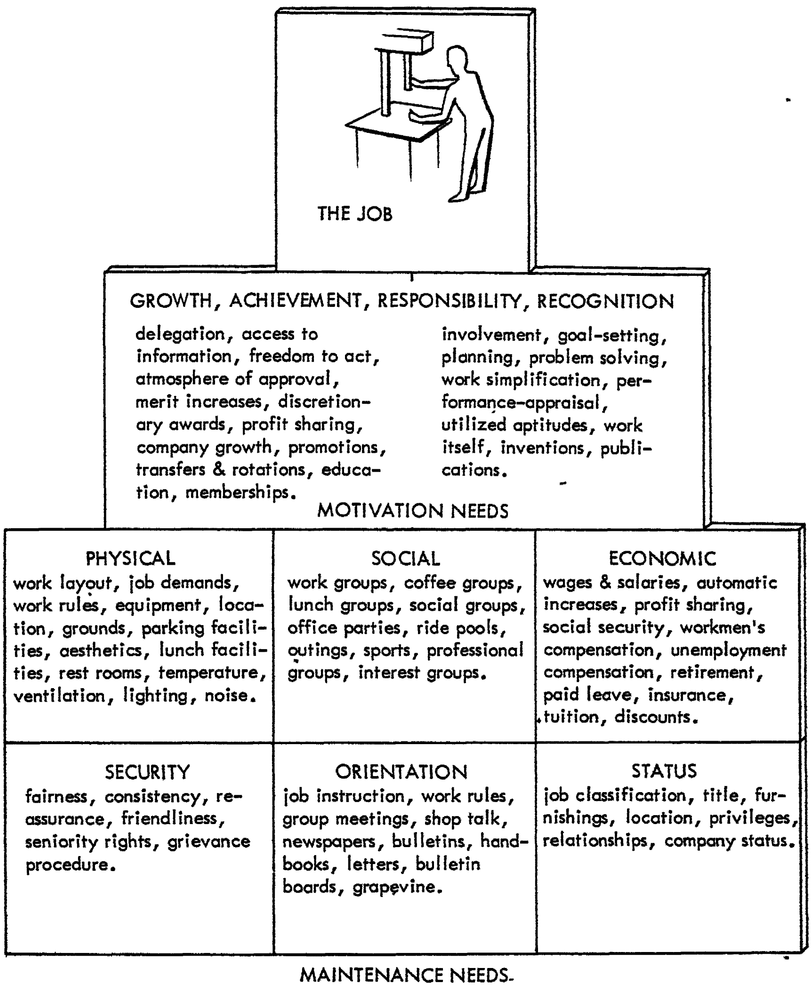


Fig. 12.2. Illustration of the motivation and maintenance needs of employees. (Adapted from Myers, Exhibit XIII.)

---

[22] *Ibid.*

on the motivation needs, it focused attention on some of the factors shown in the top half of the figure; and in its attack on the maintenance needs it concentrated on some of the environmental and work-situation factors shown in the bottom half.

INDIVIDUALITY OF MOTIVES

It is manifest that individual motivation is complex. The discussion above deals with some of the factors that have been proposed as determinants of motivation in work—hierarchies of needs, frustration, internal conflict, cognitive dissonance, attitudes, and the factors in the job and in the work situation studied by Herzberg.[23] It is not now possible to synthesize available knowledge about human motivation into a solid set of explanatory principles. It is reasonable to believe, however, that some of the factors mentioned above interact in some manner with many other factors (personal experiences, emotions, likes, dislikes, fears, etc.) to form the motivational base for each individual. Each person also has his own sets of values relating to incentives that are different from those of other people. This, in turn, leads to our difficulties in interpreting the actions of other people, since *we* have values different from *theirs,* and therefore their behavior does not make sense from *our* point of view. Many of the misunderstandings in industry can be attributed to the fact that people differ in their sets of values or needs. Management representatives, for example, may not be able to understand the motivation that induces employees to join a union. Or an employee may not understand why the foreman insists on having a particular job done on time.

FACTORS IN JOB SATISFACTION

The satisfaction which people experience in their jobs is in large part the consequence of the extent to which the various aspects of their work situations tend to be relevant to their job-related value systems. In this connection it would be useful for us to discuss a few of the many job satisfaction surveys that have been carried out, both to gain an impression of the level of job satisfaction of people, and to see what value systems people have in relation to their work.

Job Satisfaction in General Something of an overview of the job satisfaction of people is reported by Blauner.[24] On the basis of many

[23] Herzberg *et al*, *op. cit.*

[24] R. Blauner, "Extent of Satisfaction: A Review of General Research," in Costello and Zalkind, *Psychology in Administration* (Englewood Cliffs, N J.: Prentice-Hall, Inc., 1963), Chap. 6.

different surveys, he reports that, in general, about 13 per cent of the employees covered indicate that they are "dissatisfied" with their work. From these data he concludes that even under existing conditions (which are far from satisfactory) most people tend to like their jobs. Even adjusting for some tendency to exaggerate the degree of actual satisfaction, he estimates that over 80 per cent indicate general job satisfaction.

Occupational Differences in Job Satisfaction While these general findings indicate a reasonably high level of job satisfaction *generally,* there are noticeable *differences* in attitudes among various occupational groups. This was revealed, for example, by a fairly classic study by Hoppock,[25] who carried out a survey of 309 people in one community. These people were asked to fill out a questionnaire dealing largely with certain aspects of job satisfaction. Responses to certain questions are given in Table 12.2.

Table 12.2

RESPONSES TO CERTAIN QUESTIONS ON JOB SATISFACTION SURVEY

Question and response	*Percentage*
Choose the ONE of the following statements which best tells how well you like your job:	
1. I hate it	2
2. I dislike it	2
3. I don't like it	11
4. I am indifferent to it	9
5. I like it	63
6. I am enthusiastic about it	9
7. I love it	5
Check one of the following to show HOW MUCH OF THE TIME you feel satisfied with your job·	
8. All of the time	41
9. Most of the time	27
10. A good deal of the time	8
11. About half of the time	9
12. Occasionally	5
13. Seldom	5
14. Never	5
If you could have your choice of all the jobs in the world, which would you choose?	
22. Your present job	48
23. Another job in the same occupation	16
24. A job in another occupation	36
Which gives you more satisfaction?	
32. Your job	66
33. The things you do in your spare time	34

Source: *Hoppock,* op. cit., *pp. 250–252.*

[25] R. Hoppock, *Job Satisfaction* (New York: Harper & Row, Publishers, 1935).

It can be seen that people differ in the degree to which their jobs are satisfying to them. It might be pointed out, however, that many of the people covered by this survey indicated that they liked their jobs. The degree of job satisfaction is somewhat related to the type of job, as shown in Table 12.3. This gives the range, and the mean, of the "job satisfaction indexes" of people in each of various occupational classifications. The job satisfaction index, derived from responses to the questionnaire, can range from 100 (minimum satisfaction) to 700 (maximum satisfaction). Although the mean indexes do differ for the various occupational groups, it can be seen that the ranges for the groups are all fairly wide, indicating that job satisfaction, even within each occupational group, varies considerably from individual to individual.

Table 12.3

JOB SATISFACTION INDEXES OF FIVE OCCUPATIONAL GROUPS

Occupational group	*No. of cases*	*Range of indexes*	*Mean index*
Unskilled manual	55	100–650	401
Semiskilled	74	125–650	483
Skilled manual and white-collar	84	125–675	510
Subprofessional, business, and minor supervision	32	250–700	548
Professional, managerial, and executive	23	300–700	560

Source: *Hoppock,* op cit, *p. 225.*

Another investigation is reported by England and Stein.[26] These investigators administered an Employee Attitude Questionnaire to 3,207 non-supervisory employees in 26 different companies. The survey covered 11 areas of job satisfaction, namely:

1. Working conditions
2. Company
3. Pay
4. Hours
5. Co-workers
6. Type of work
7. Supervision
8. Promotions
9. Communications

---

[26] G. W. England and C. I. Stein, "The Occupational Preference Group—A Neglected Concept in Employee Attitude Studies," *Personnel Psychology,* 14 (1961), 299–304.

10. Recognition
11. Security

The data obtained were analyzed separately for seven occupational groups as follows:

1. Professional
2. Sales
3. Clerical
4. Craft
5. Semiskilled
6. Labor
7. Service

The results of this survey showed significant differences among the various occupational groups in terms of the per cent answering "strongly agree" to questionnaire items in the various areas. These differences led the investigators to argue for the use of special attitude scales for different occupational groups, and for the use of occupational norms when analyzing such data. It is interesting to note from a further analysis of the data, however, that the median rank order correlation between rank of importance of the attitude areas for all of the occupational groups was .72. This rank order indicates at least a moderate agreement in terms of *relative* importance of the various attitude areas. Such a correlation, however, does not reflect the differences in the *absolute* levels of attitudes in the various areas. Many of these levels were significantly different.

Factors Associated with Occupational Differences The factors that account for occupational differences in satisfaction are not readily identifiable. In this connection, however, Blauner [27] has attempted to explain the gross differences that exist among people in different occupations and industries, and has set forth four factors that seem useful in explaining the differences. These are: (1) occupational prestige; (2) control; (3) integrated work groups; and (4) occupational communities.

Of these, differences in occupational prestige seem to be particularly important, as reflected by the fact that the rank order of job satisfaction of various occupational groups corresponds generally with the rank order of prestige of the groups. The control factor deals with the relative amount of "control" inherent in jobs. Satisfaction generally is higher in the case of people whose jobs involve control over their own work and that of others, and is lowest for those people who are in jobs that are at the lower end of the organizational hierarchy, for whom there is little opportunity for such control. The third factor relates to the degree of integration of work groups. In general, the higher the degree of integration of work groups, the higher the level of job satisfaction. The de-

[27] Blauner, *op. cit.*

gree of integration usually is a function of the nature of the work, and the requirement it places upon people to work together as a team. The fourth factor, occupational communities, refers to the nature of the association among workers off-the-job. One characteristic of such association that is cited by Blauner [28] consists of isolated communities of people who generally are engaged in the same work; an example is that of mining, which typically requires that people live together in a community made up largely of fellow workers. On the other hand, he points out that there is relatively little such off-the-job socialization among urban factory workers. Another characteristic of an occupational community is that its participants "talk shop" in their off hours, such as in the case of farmers, fishermen, miners, and railroaders. Still another characteristic of occupational communities is that they are little worlds in themselves; the occupational group is a reference group in that its standards, and systems of status and rank, guide the conduct of the group, such as in railroading.

While it is, of course, not feasible to modify all of these factors in jobs generally, there are certain kinds of action which management can take in some situations to modify at least certain ones, toward the end of enhancing the opportunity for greater job satisfaction on the part of workers. For example, in certain situations steps could be taken to provide greater opportunity for individual "control" of certain aspects of their jobs, or to organize the work so that people become members of work teams, or to provide for greater job prestige by appropriate means.

What People Want from Their Jobs Job satisfaction is, of course, basically an individual matter. Individuals look for those aspects of jobs which are related to their own value systems, some placing greater value on, say, security, others on income, others on the type of work, etc. But, while individuals vary in their sets of values, there sometimes is at least moderate consistency in the values of individuals within specific groups. It should be added, incidentally, that the values are subject to at least moderate change depending upon economic, social, and other influences. An understanding of the value systems of employees can be very useful to management in developing personnel programs. If, for example, management operates on the philosophy that all that the employees are interested in is money, it might be surprised to find that, despite high wages, the employees vote to go out on a strike over some apparently trivial matter.

In one part of the study by Hoppock,[29] 80 persons representing a wide range in age, intelligence, occupations, and earnings were interviewed. The interview case studies corroborate the general conclusion

[28] *Ibid.*
[29] Hoppock, *op cit.*

reached above—that is, that job satisfaction is related to a good many things besides financial return. Some of these factors, as summarized by Hoppock, are "relative status of the individual within the social and economic group with which he identifies himself, relations with superiors and associates on the job, nature of the work, earnings, hours of work, opportunities for advancement, variety, freedom from close supervision, visible results, the satisfaction of doing good work, opportunities for service to others, environment, freedom to live where one chooses, responsibility, vacations, excitement, opportunity for self-expression, competition, religion, opportunity for or necessity of travelling, fatigue, appreciation of criticism, security, and ability to adjust oneself to unpleasant circumstances."

A study by Gruenfeld[30] shows the relative importance of 18 job characteristics to typical industrial supervisors. Gruenfeld had 52 industrial supervisors pair-compare 18 job characteristics. The resulting rank order of the 18 job characteristics is shown in Table 12.4.

Table 12.4

RANK ORDER OF
18 JOB CHARACTERISTICS
AS JUDGED BY
18 INDUSTRIAL SUPERVISORS

Rank	*Job Characteristics*
1.	Greater opportunity for advancement (promotion)
2.	Better opportunity for education and self-development
3.	More opportunity to see concrete results of my work
4.	A higher degree of personal responsibility
5.	More opportunity for independent action
6.	More opportunity to lead and develop subordinates
7.	Greater job security
8.	More opportunity for close association with higher-level management
9.	Higher wages
10.	More prestige within the company
11.	More frequent and closer contact with workers
12.	More power and authority
13.	Fewer people to please, less criticism
14.	Fewer worries, tensions, and troubles
15.	Better fringe benefits
16.	Less need for dealing directly with workers
17.	More definite and regular working hours
18.	Safer, cleaner, less fatiguing work

In Gruenfeld's study of industrial supervisors, higher wages was found to be ninth among the 18 job characteristics investigated. At the

[30] L. W. Gruenfeld, "A Study of the Motivation of Industrial Supervisors," *Personnel Psychology*, 15 (1962), 303–314.

top of the list were found greater opportunity for advancement and better opportunity for education and self-development.

The results of several different surveys of this type are summarized in the first seven columns of Table 12.5 (the columns headed "Relative importance of factors"). For each group surveyed, the table shows the average rank order of importance of each factor that was ranked by the individuals in the group; (some factors were not included in all surveys, and the terms used in some cases were slightly different from those given in the table). While there is a reasonable amount of similarity in rank orders across the several groups, there are certain noticeable differences. As a matter of interest, it should be noted that, in general, opportunities for advancement and job security were high on most lists, and pay was generally around the middle.

The last column shows the rank order of *actual* satisfaction as perceived by the people covered by the survey by England and Stein.[31] (It will be recalled that this survey covered 3,207 nonsupervisory employees in 26 companies, and included seven occupational groups. The rank orders given are those of all groups combined.) The data given reflect the rank order of *actual* satisfaction with the different factors, that is, the relative extent to which the employees perceived the different factors to be fulfilled in their jobs. These data are presented here for comparative purposes, to reflect the disparity between *actual* satisfaction (from this study) as compared with rank order of *importance* (as reflected by the others). Some of these differences are of considerable magnitude, and point up the fact that some factors were not being adequately fulfilled in the work situations of the employees in question.

In interpreting the results of such surveys, one should keep in mind that they reflect averages of the responses of many individuals. While such averages are, of course, meaningful in an overall sense, there still can be differences in the value systems of the individuals who are included.

In considering such data as those given in Table 12.5, one might wonder how sensitive management officials and union leaders are to the attitudes of employees regarding the relative importance of various job factors. Some indications regarding this come from a survey by Raube.[32] He submitted a list of 71 morale factors to employees of six companies, 50 executives of these companies, and 42 labor leaders. Each employee was asked to identify in order of importance the five factors that were most important to him. The executives were asked to predict the rankings that would be assigned to these factors by the employees of their respective companies, and the labor leaders indicated the rankings that they believed represented the relative importance of the factors to their union

[31] England and Stein, *op. cit.*

[32] S. A. Raube, "Factors Affecting Employee Morale," *Studies in Personnel Policy, No. 85* (New York: National Industrial Conference Board, 1947).

Table 12.5

RANK ORDER OF JOB FACTORS AS OBTAINED FROM VARIOUS STUDIES

	Relative importance of factors							Level of actual satisfaction
Source	*Chant*	*Chant*	*Wyatt et al.*	*Berdie*	*Blum and Russ*		*Jurgensen*	*England and Stein*
Group	*Misc. workers*	*Dept. store workers*	*Women factory workers*	*Male H.S. graduates*	*Male*	*Female*	*Male applicants*	*Male employees*
Number	150	100	325	150	181	105	3345	3207
Opportunity for advancement	1	1	5	2	1	1	2	11
Job security	2	2	1	1	2	2	1	5.5
Opportunity to use ideas	3	3	7	4				
Opportunity to learn a job	4	4	8	7				
Opportunity for public service	5	7		8				
Type of work							3	2
Supervisor	6	5	4	9	4	3	7	3
Company							4	7
Pay	7	6	6	3	3	4	5.5	10
Co-workers	8	8	3	5			5.5	1
Working conditions	9	9	2	11			9	8
Clean work	10	11		10				
Working hours	11	10	9	6	5	5	8	4
Easy work	12	12	10	12				
Benefits							10	
Communications								5.5
Recognition								9

Note: Entries of 5.5 represent ties for fifth and sixth ranks.
S. M. F. Chant, "Measuring the Factors that Make a Job Interesting," Personnel Journal, *11 (1932) 1–4.*
S. Wyatt, J. N. Langdon, and F. G. L. Stock, "Fatigue and Boredom in Repetitive Work," Industrial Health Research Board, Report No. 77 *(London, 1937).*
R. F. Berdie, "Can Factors in Vocational Choice be Weighted?," Occupations, *22 (1943), 43–46.*
M. L. Blum and J. J. Russ, "A Study of Employee Attitude Toward Various Incentives," Personnel, *19 (1942), 438–444.*
C. E. Jurgensen, "What Job Applicants Look for in a Company," Personnel Psychology, *1 (1948), 443–445.*
England and Stein, op. cit.

members. Definite differences in the rankings were found for the three groups. The employees attached most importance to the following five:

1. Job security.
2. Compensation.
3. Opportunities for advancement.
4. Employee financial benefits (insurance, and so forth).
5. Informing employees about their job status.

Executive predictions agreed with the employee's ranking in only three of the most important five (job security, compensation, and opportunities for advancement), while labor leaders agreed with the employees in only two (job security and compensation). While compensation and job security were selected as two of the five most important factors by all three groups, there was little agreement in identifying the other factors considered by the employees to be most important. The results of this comparison suggest that when management is interested in knowing how employees feel about various aspects of the work situation, it is desirable to find out directly (such as by the use of attitude surveys) rather than trying to make their own judgments on such matters.

EMPLOYEE ATTITUDES AND BEHAVIOR

As indicated in the early part of this chapter, attitudes have motivational origins. The results of various types of attitude surveys, including those relating to job satisfaction, then have some relevance to employee motivations. In this connection, it has been generally assumed that employees with favorable attitudes are in general "better" employees (in terms of productivity, job tenure, and other criteria) than are employees with less favorable attitudes. Because of the importance of this question, we should examine this assumption.

In a comprehensive survey of studies relating to employee attitudes, Brayfield and Crockett[33] were especially interested in crystallizing, if they could, any general implications of the many studies regarding the relationship between general employee attitudes and "performance." Various aspects of employee "performance" were considered, including the following: job performance as such, absences, tardiness, and tenure (referred to as "employment stability").

Employee Attitudes and Job Performance It must be stated that the survey brought out information that is somewhat at odds with common beliefs or assumptions. In particular there seemed to be no consistent pattern of relationship between attitudes and opinions on the one hand and actual job performance on the other hand.

---

[33] A. H. Brayfield and W. H. Crockett, "Employee Attitudes and Performance," *Psychological Bulletin*, 52 (1955), 396–428.

Two or three studies will be discussed in order to illustrate the divergencies of the results of such studies. In one study Mossin [34] investigated the sales performance of 94 female salesclerks in a department store in relation to the salesclerks' scores on a job satisfaction questionnaire. Sales performance was quantified by ratings. The correlation between job satisfaction scores and ratings on sales performance was −.03, indicating the absence of any significant relationship.

Another study with somewhat similar results was reported by Bernberg [35] in an aircraft plant. He used four attitude measures that need not be described in detail; briefly, they measured "group morale," the employee's "acceptance" of the formal organization, attitude toward supervisor, and "self-rating" scale. These were correlated with job performance ratings, the correlations ranging from .02 to .05, which are almost as close to zero as correlations can get.

An illustration of a study with somewhat different results is one that was carried out at The Prudential Life Insurance Company by The Survey Research Center of the University of Michigan.[36] In this study it was possible to identify certain "high" producing sections and other "low" producing sections, since various sections were doing identical types of office work, and since there were records available of the actual clerical time spent in completing a given amount of work. In one phase of the study four attitudinal variables were developed, namely: (a) pride in work; (b) intrinsic job satisfaction; (c) involvement in the company; and (d) financial and job status. Attitudes on these four variables were obtained by the use of interviews. After deriving attitude scores on these variables from the interview, it was possible to compare the attitudes of employees in the "high" and "low" sections.

The results are summarized in Fig. 12.3. This gives, for the "high" and "low" producing sections, the percentages who had high, medium, and low "morale" in each of the four attitude areas. It was found that "pride in work" bore a significant relationship to the work performance of the various sections, but that the other attitude areas did not.

The primary implication of the rather diverse studies relating to attitudes and job performance is that satisfaction with one's work or work situation apparently does *not necessarily* provide strong motivation to high levels of job performance. In fact, Kahn [37] asserts, without qualifi-

---

[34] A. C. Mossin, *Selling Performance and Contentment in Relation to School Background*, Bureau of Publications, Teachers College, Columbia University, 1949.

[35] R. E. Bernberg, "Social-Psychological Factors in Industrial Morale: I. The Predictions of Specific Indicators," *Journal of Social Psychology*, 36 (1952), 73–82.

[36] D. Katz, N. MacCoby, and Nancy C. Morse, *Productivity, Supervision and Morale in an Office Situation*, Survey Research Center, Institute for Social Research, University of Michigan, December, 1950.

[37] R. L. Kahn, "Productivity and Job Satisfaction," *Personnel Psychology*, 13 (1960), 275–287.

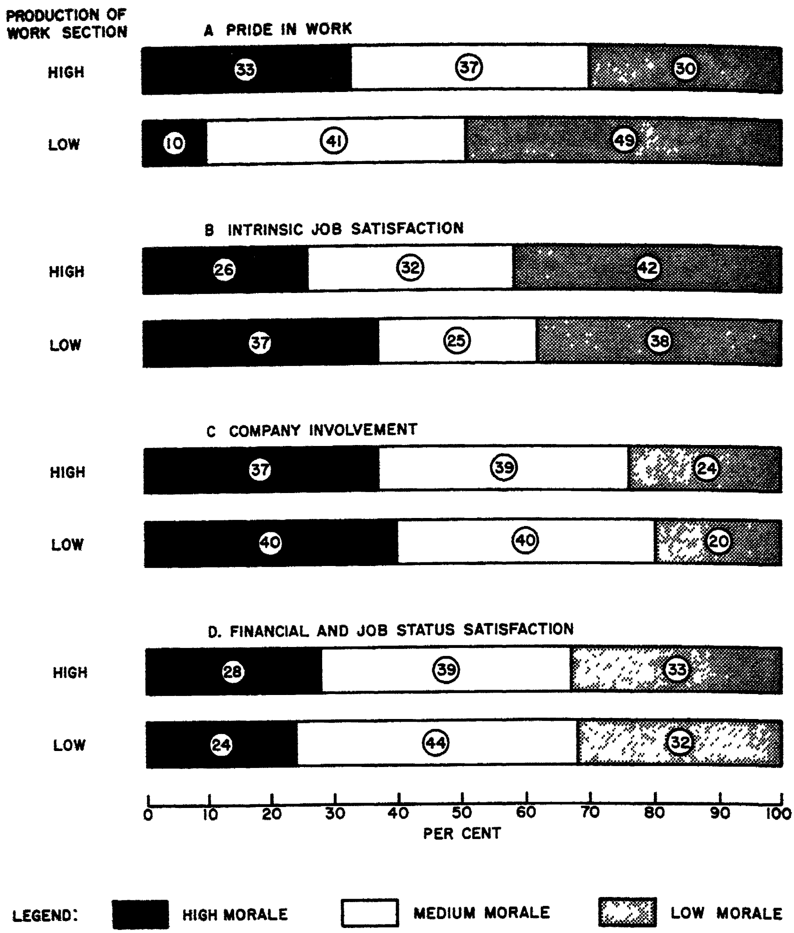


Fig. 12.3. Per cent of employees in high and low producing sections who had high, medium, and low "morale" indexes for each of four morale areas. The relationship between "morale" and section productivity was significant only for the first morale area, namely, "Pride in Work" groups. (Adapted from Katz, MacCoby, and Morse, *op. cit.*, Tables 23, 25, 26, and 32.)

cation, that productivity and job satisfaction do not necessarily go together. The fact that some individual studies have revealed positive relationships between attitudes and productivity, however, suggests that there presumably are certain *kinds* of circumstances in which job satisfaction is related to work performance, and *other* kinds of circumstances where this is *not* the case. So far the variables that account for these differences have not been adequately identified. As pointed out by

Wilensky,[38] there may be a number of such variables, such as: economic and technological (employment levels, business cycle, and the like); social and cultural background of employees; nature of work incentives; or the nature of the leadership in the organization.

Employee Attitudes and Absenteeism, Tardiness, and Tenure Although attitudes do not seem consistently to be related to job performance as such, there is somewhat more consistency in the relationship of attitudes to such aspects of behavior as absenteeism, tardiness, and tenure or "employment stability." Even with these criteria the reported studies do not indicate strongly consistent patterns. but the various pertinent studies suggest at least a moderate (though complex) relationship between attitudes and absenteeism, tardiness, and employment stability. A few examples will be mentioned briefly.

In a study that dealt primarily with attitudes toward merit rating, Van Zelst and Kerr[39] also obtained some information pertinent to our present subject. A total of 340 employees in 14 companies completed a questionnaire that included their own reports of absenteeism and tardiness. In addition, the questionnaire included two questions regarding job satisfaction. Responses to these two questions were combined to give a single index. This job satisfaction index had a correlation with a reported favorable absentee record of .31, and with a reported favorable tardiness record of .26. Both correlations were statistically significant.

In a study by Giese and Ruter,[40] an analysis was made of the average "morale index" of employees in 25 different departments of a mail-order house in relation to tardiness and absence indices of those departments. The average morale index had no significant relationship with tardiness, but had a statistically significant correlation with absences of —.47 (this can be interpreted as meaning that departments with *most* absences tended to have *low* morale, and vice versa).

In still another absenteeism study, Mann and Baumgartel[41] took a look at various aspects of attitudes and opinions of both white-collar and blue-collar male employees of an electric power company. In one aspect of the study, "over-all satisfaction" of white-collar employees was obtained by the use of a questionnaire. This index was then compared with absenteeism rates for the various work groups (departments, units, and so

---

[38] H. Wilensky, "Human Relations in the Workplace: An Appraisal of Some Recent Research," in *Research in Industrial Human Relations* (New York: Harper & Row, Publishers, 1957).

[39] R. H. Van Zelst and W. A. Kerr, "Workers' Attitudes Toward Merit Rating," *Personnel Psychology*, 6 (1953), 159–172.

[40] W. J. Giese and W. H. Ruter, "An Objective Analysis of Morale," *Journal of Applied Psychology*, 33 (1949), 421–427.

[41] F. Mann and H. Baumgartel, *Absences and Employee Attitudes in an Electric Power Company*, Survey Research Center, University of Michigan, December, 1952.

forth). The results (shown in Fig. 12.4) show that, of the employees in groups with absence indices of 4 or more, only 22 per cent were "satisfied" with their jobs. By comparison, in groups with absence indices of 1, and of 2 or 3, the percentages of employees who were "satisfied" were 62 and 52,[42] respectively.

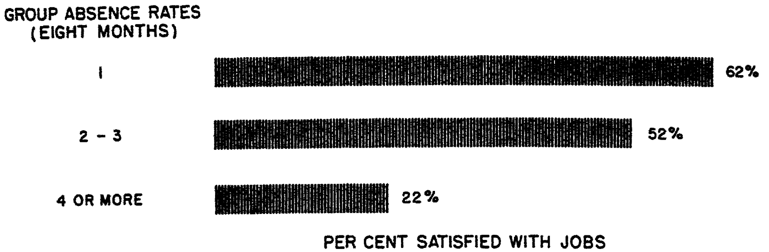


Fig. 12.4. Relationship between over-all satisfaction with company and job, and absence rate, of male white-collar employees. (Adapted from Mann and Baumgartel, *op. cit.*, p. 22.)

In connection with tenure (or its corollary of turnover) and employee attitudes, Weitz and Nuckols,[43] report a statistically significant correlation of .20 between attitudes of 1,200 insurance agents and their tenure. Kerr [44] reports a significant correlation of .25 for a sample of 98 wage-earners between their "job satisfaction" scores and their self-reported tenure (number of years in the labor market divided by number of employers). In turn, Giese and Ruter [45] report a correlation of "morale scores" with a turnover criterion of −.42 for 25 departments. This *negative* correlation, based on *turnover indices,* of course indicates a *positive* relationship with *tenure.*

Although some other studies have revealed negligible relationship between employee attitudes and tenure, there nonetheless are a number of situations in which there is some relationship between attitudes and tenure, although the relationship is not very great. This also seems to be true of absenteeism and tardiness.

Comments In general, then, we cannot say unequivocally that employees with favorable general attitudes are necessarily more *productive* than those with unfavorable attitudes, although there are some situations

[42] The 52 per cent is an average of the percentages shown in the original data for employees in groups with indices of 2 and 3; those percentages were 49 and 56, respectively.

[43] J. Weitz and R. C. Nuckols, "The Validity of Direct and Indirect Questions in Measuring Job Satisfaction," *Personnel Psychology,* 5 (1953), 487–494.

[44] W. A. Kerr, "On the Validity and Reliability of the Job Satisfaction Tear Ballot," *Journal of Applied Psychology,* 32 (1948), 275–281.

[45] Giese and Ruter, *op. cit.*

where this has been found to be the case. There does seem to be some support for saying that attitudes are moderately related to other aspects of employee behavior such as absenteeism, tardiness, and tenure. To maintain an objective, scientific attitude toward this matter, it must be pointed out that the *fact* of this relationship does not *necessarily* imply *cause and effect*. It may be, for example, that those employees who, generally speaking, tend to have favorable attitudes toward their employment situation also are the kinds of people who tend to come to work more regularly, to be on time, and to stay longer on their jobs. In other words, from the strictly scientific point of view, one *cannot* say *if* a company is able in some way to change the attitudes of its employees, that the employees, as a *consequence*, will improve in those aspects of their behavior. To do this requires that one make a basic assumption, namely, that there is a cause-and-effect relationship between employee attitudes and employee behavior (such as absenteeism, tardiness, and tenure, and possibly job performance and other aspects of behavior). This requires that one shift gears from a completely objective slant to one in which we must rely on subjective evaluations, value judgments, hints from experiments, and "logic," with perhaps a dash of rationalization thrown in.

Taking such liberties in assessing the potpourri of information regarding human behavior, especially in an industrial setting, the authors feel that there is some justification for believing that the attitudes of employees do have some effect on employee behavior—or, to put it another way, there will be benefits both to employees and to the organization if the organization is able to create those conditions that are conducive to the formulation of favorable employee attitudes.

It may be, however, that the values to the organization are difficult to identify or measure. There may be values *other* than those having a direct influence on employee productivity (or for that matter, absenteeism, tardiness, and tenure). Brayfield and Crockett,[46] for example, suggest that conditions conducive to job satisfaction may have an effect on the quality of the applicants drawn into the organization, on the quality of job performance, and on the harmony of labor-management relations. There may be various types of "hidden" costs in having dissatisfied employees—costs such as having to have more supervision, greater time devoted to handling grievances and complaints, poor community relations and unmeasurable (but nonetheless real) effects on the total organizational efficiency.

Our discussion of employee attitudes so far has been from the point of view of the employing organization. The point of view of employees themselves also should be emphasized. The total welfare of individuals, and of society as a whole, depends, in part, on the satisfactions that people

---

[46] Brayfield and Crockett, *op. cit.*

experience in the various aspects of their lives, including their employment relationships. In recent years there has been increasing acceptance on the part of industry of "social" obligations to the communities and to the nation as a whole.

Thus, in terms of both the immediate interests of industry itself and of long-range human welfare, there seems to be adequate justification for the actions taken by management to create work situations that are conducive to the increase of human satisfactions.

13

Human Behavior in Organizations

In Chapter 12 we discussed some of the general aspects of human motivation. Still keeping in mind that behavior has its origins in motivation with specific incentives serving as foci of goal-directed behavior, let us now turn our attention to some of the implications of human behavior in industrial organizations.

ORGANIZATIONS AND PEOPLE

One of the most persistent and perplexing problems in the management of organizations is that of achieving reasonably optimum use of the talents of individuals in the organization. This is a two-sided coin in that to achieve this objective, two conditions must obtain. In the first place, the individual must be willing to apply his talents; he must be motivated. And, in the second place, the characteristics of the organization (its structure, policies, etc.) must be such as to encourage such motivation and to facilitate the the efficient use of personnel in working toward its objectives.

The central question for the management of an organization, however, concerns *what* characteristics of organizations facilitate personnel motivation and operating effectiveness. Beyond this, there is the admin-

istrative problem of bringing these characteristics into being within the organization.

In very general terms, these characteristics include organizational structures, managerial principles and practices, communications, personnel, and, perhaps most important, the prevailing atmosphere of the organization. It is to some of these considerations that we will now turn our attention. In particular, for each of several such considerations, we will examine some empirical evidence relating to the relationship between pertinent situational variables and human behavior, and will discuss related principles and theories of behavior. Where feasible we will also offer some suggestions regarding ways and means of implementing those conditions that tend to be conducive to effective behavior in work situations.

Management Intention The introduction of organizational changes that are intended to bring about improved employee behavior, however, raises justifiable apprehensions about possible manipulation and exploitation, as indicated by McGregor.[1] In this connection, a statement by Worthy[2] is uniquely appropriate:

> I agree . . . that gimmicks and devices employed for purposes of manipulation will soon lose their effectiveness. The important question is management's motives in employing the results of human relations research. If its motives are those of narrow self interest, of finding subtler and smoother ways of bending workers to its will, the effort will be worse than useless for it will widen further the gap between workers and management. But if management's motives are sincerely those of better understanding the problems of people at work, of finding ways for making work a more rewarding experience, of discovering its own shortcomings and means for improvement, management's efforts to apply the findings of human relations research are likely to create positive benefits for all concerned.

Human Behavior and Incentives in Industry In considering human behavior in industry, it is in order to mention specifically some of the incentives that are operative in the industrial setting. (Let us keep in mind, however, that their relative values for individuals will vary markedly.) The incentives can be either formal or informal, and either positive or negative. Some examples of these are given below:

Formal incentives

Positive: money, bonuses, promotions, awards, formal commendations, special privileges (membership in clubs, officer dining room privilege, special parking location, etc.), choice of work schedule.

Negative: reprimands, disciplinary actions, demotions, lay-offs, discharges, withholding or withdrawal of privileges.

---

1 D. McGregor, *The Human Side of Enterprise* (New York: McGraw-Hill Book Company, 1960), p. 12.

2 J. C. Worthy, "Comments on Mr. Wilensky's Chapter," in *Research in Industrial Human Relations* (New York: Harper & Row, Publishers, 1957).

Informal incentives

Positive: praise, encouragement, friendly attitudes by others, acceptance by group, minimum supervision, respect by management and fellow workers.

Negative: disapproval and rejection by others, criticism, assignment to more onerous job, lack of work cooperation by fellows, being "picked on" by supervisors and fellow workers.

MANAGEMENT THEORIES AND PHILOSOPHIES

From the welter of behavioral research in industrial settings, it has been clearly indicated that top management typically is the most critical factor in creating the psychological "atmosphere" of an organization—for better or for worse, as the case may be. The influence of the patterns of behavior and attitudes on the part of top management tend to set off something like a chain reaction through the levels of the organization. This was illustrated, for example, by the results of a study by Fleishman [3] that involved leadership training for foremen, the training being away from the company. While there was evidence of change on the part of the foremen by the completion of their training, when they returned to their jobs they tended to adopt the behavior patterns of *their* superiors. If the supervisor of a foreman generally behaved on the basis of different leadership principles than those covered in the training program, the foreman, when back on the job, would tend to pattern his behavior more like that of *his* supervisor than that covered by the training itself.

This is explainable on the basis of the motivational effects of the rewards and punishments that an individual receives in doing his job. In the above situation, a foreman presumably would be "rewarded" for behavior of which his superior approved, and "punished" otherwise; thus, the behavior and attitudes of subordinates tend to be molded somewhat along the lines of the behavior and attitudes of *their* superiors via the powerful forces of rewards and punishment.

Opposing Theories of Management As pointed out by McGregor,[4] virtually every act on the part of management and supervisory personnel that involves human beings is predicated upon some assumptions, generalizations, and hypotheses relating to human nature and human behavior (in other words on some theory of behavior). These assumptions may be neither consciously crystallized nor overtly stated, but they none the less serve to influence our predictions about human behavior. Thus, if a subordinate makes a serious error, and the supervisor patiently explains again to him the correct procedures to be followed, he is

[3] A. F. Fleishman, "Leadership Climate, Human Relations Training, and Supervisory Behavior," *Personnel Psychology*, 6 (1953), 205–222.

[4] McGregor, *op. cit.*, p. 6.

operating on the assumption that the further information to the subordinate will be more effective in avoiding such future events than a reprimand would have been.

The basic assumptions that different people have regarding human behavior in industry vary considerably, but there are two generally obverse points of view that have been characterized. These have been referred to by McGregor [5] as theory X and theory Y.

Theory X: the traditional view. The traditional view of human behavior that has been accepted by some companies is characterized by certain assumptions, including the following: [6]

1. The average human being has an inherent dislike of work and will avoid it if he can.
2. Because of this, most people must be coerced, controlled, directed, and threatened with punishment to get them to put forth adequate effort toward achievement of the organizational objectives.
3. The average human being prefers to be directed, wishes to avoid responsibility, has relatively little ambition, wants security above all.

In accentuated form, this theory of behavior is characteristic of organizations that specify very rigid standards of work behavior, and have stringent rules and regulations that are rigorously enforced. In more modified form some companies have adopted what are purported to be new approaches to human relations, although (as pointed out by McGregor) some of these are simply old wine in new bottles; they are different tactics (programs, procedures, gadgets, etc.) based on the same, unchanged strategy of theory X.

It should be noted that some human behavior probably can be explained in terms of this theory; the underlying assumptions probably would not have persisted had there not been some confirmation of it in the practical affairs of industry. But, on the other hand, there are many aspects of human behavior that are incompatible with such a theory.

Theory Y: the integration of goals. The other theory postulated by McGregor [7] is predicated upon such assumptions as the following:

1. The expenditure of physical and mental effort in work is as natural as play or rest.
2. External control and the threat of punishment are not the only means for bringing about effort toward organizational objectives. Man will exercise self-direction and self-control in the service of objectives to which he is committed.
3. Commitment to objectives is a function of the rewards associated with their achievement.

---

[5] McGregor, *op. cit.*, Chapters 3 and 4.
[6] McGregor, *op. cit.*, pp. 3–4.
[7] McGregor, *op. cit.*, pp. 47–48.

4. The average human being learns, under proper conditions, not only to accept but to seek responsibility.
5. The capacity to exercise a relatively high degree of imagination, ingenuity, and creativity in the solution of organizational problems is widely, not narrowly, distributed in the population.
6. Under the conditions of modern industrial life, the intellectual potentialities of the average human being are only partially utilized.

The core of this theory is that of integration of goals of individuals and of organizations—of creating those conditions in which the members of an organization can best achieve their goals by directing *their* efforts toward achievement of the goals of the organization.

Discussion To some degree, both of these theories probably are self-confirming. If management expects people to act in line with theory X, and treats the employees accordingly, the employees may very well behave in somewhat the expected manner. Similarly, if management treats employees in the manner implied by theory Y, the employees may behave in the manner so implied.

These two divergent points of view (characterized by McGregor [8] as theory X and theory Y) have been variously described, labelled, and discussed, by others. Likert,[9] for example, refers to *job-organization* and *cooperative-motivation* systems. The job-organization system is the expression of the traditional management philosophy, as implied in theory X above, with a strong emphasis on the economic motives of "buying" a man's time, and telling him precisely what to do and how to do it. As Likert points out, this system has had considerable success, especially in industries that involve a great deal of repetitive work. The cooperative-motivation system, which corresponds essentially to theory Y, places a greater emphasis on the ego motives. It has been used more in situations where varied work is more dominant. In such situations there is greater opportunity for the development of enthusiasm regarding the job itself, and the achievements of the job.

While there are many unanswered questions about such opposing theories, there is a gradually accumulating body of evidence and opinion to support some of the postulates of theory Y (or the cooperative-motivation system). Some of those who have studied this area intensively have adopted very strong positions. Katz,[10] for example, proposes what he admits is a heretical proposition to the effect that the conventional way of thinking about how an enterprise should be organized and administered is obsolete (such concepts relate to emphasis on profits, policy

---

[8] McGregor, *op. cit.*

[9] R. Likert, *New Patterns of Management* (New York: McGraw-Hill Book Company, 1961).

[10] R. L. Katz, "Toward a More Effective Enterprise," *Harvard Business Review* (September–October, 1960).

planning, program formation, and delegation of responsibility to successively lower levels). Such concepts typically result in management's devising a predetermined program for reaching goals, in "ordering" employees to carry out such programs, in administering rewards and punishments, and in judging accountability for results.

As possible alternatives to these conventional management concepts, Katz[11] proposes certain alternative assumptions. One of these would involve shifting the focus from profits per se, to the *tasks* to be performed —the work activities of individuals—in order to provide greater opportunity for those involved to satisfy their needs. Such a shift would not be to deny the importance of profits, but rather would operate in the direction of integration of the goals of individuals and those of the organization. An internal focus on task activities and people and their relationships would minimize the conception of people serving as "instruments" for profit as such. Incidentally, he cites a couple of studies which suggest that focusing on the persons, and the organizational satisfaction of their needs, may, in fact, bring about higher levels of productivity than when productivity is the sole concern of management. Other alternatives suggested by Katz include: greater emphasis in the evaluation of organizational success on the extent to which the satisfactions of members are fulfilled; greater flexibility in planning with increased latitude for possible individual action; and providing for greater contributions by individuals to organizational processes of goal-setting, planning, organizing, directing, and controlling.

An Evolving Theory Research evidence regarding human behavior in organizations does not yet make it possible to state with any degree of finality any set of principles of management that would be applicable on an across-the-board basis. A step in this direction, however, has been proposed by Likert,[12] stemming in part from some of the research carried out at the Institute for Social Research at the University of Michigan. For what interest this theory may have, some aspects of it are mentioned briefly below. It is predicated on a couple of generalizations that come from research findings:

1. The supervisors and managers in American industry and government who are achieving the highest productivity, lowest costs, least turnover and absence, and the highest levels of employee motivation and satisfaction display, on the average, a different pattern of leadership from those managers who are achieving less impressive results. The principles and practices of these high-producing managers are deviating in important ways from those called for by present-day management theories.

---

[11] *Ibid.*

[12] Likert, *op. cit.*, Chapter 8.

2. The high-producing managers whose deviations from existing theory and practice are creating improved procedures have not yet integrated their deviant principles into a theory of management. Individually, they are often clearly aware of how a particular practice of theirs differs from generally accepted methods, but the magnitude, importance, and systematic nature of the differences when the total pattern is examined do not appear to be recognized.

Analysis of research findings, however, indicates that the general operations of high-producing managers generally contrast with those of low-producing managers in some of the following respects: [13]

1. A preponderance of favorable attitudes of members of the organization toward all aspects of the job (toward other members, supervisors, the work, the organization); the favorable attitudes toward the organization and work are attitudes of identification with the organization and its objectives.
2. The highly motivated, cooperative orientation is achieved by harnessing effectively all the major motivational forces which can exercise significant influence in an organizational setting (the ego, security, curiosity, and economic motives).
3. The organization consists of a tightly knit, effectively functioning social system of interlocking work groups.
4. Measurement of organizational performance is used primarily for self-guidance rather than for superimposed control.
5. Widespread use of participation is employed in various aspects of the job and work.
6. Full advantage is taken of the technical resources of the classical theories of management, such as work simplification, budgeting, and financial controls, but these are used in different ways than by the low-producing managers; in particular these techniques are used in such a way as to bring into play some of the motives mentioned above (such as through participation) as opposed to being used to exercise control through authority.

The general principle that is the common denominator of these methods of operation is referred to by Likert as the *principle of supportive relationships,* and is stated as follows: [14] The leadership and other processes of the organization must be such as to ensure a maximum probability that, in all interactions and all relationships with the organization, each member will, in the light of his background values and expectations, view the experience as supportive and one which builds and maintains his sense of personal worth and importance. In less formal terms, this implies that the optimum conditions for organizational success are those

---

13 Adapted from Likert, *op. cit.,* Chapter 8.
14 Likert, *op. cit.,* p. 103.

in which the mission of the organization is seen by its members as being important, and where individuals perceive their own tasks as being indispensable to the achievement of the organizational objectives. It is suggested by Likert that, where jobs are not perceived as being important and meaningful in this context, they should be reorganized so that they are so perceived. While a theory postulated on the basis of some of these considerations probably cannot be accepted in its totality at present, it is reasonable to believe that continuing behavioral research will support its central features.

How Theories of Management are Implemented Whether a "theory" of management is recognized as such, managers and supervisors operate on the basis of some assumptions about human behavior. As indicated before, virtually every act of management or supervision is predicated on some assumption(s). These underlying assumptions—or theories—whether formally recognized or implicit—have their impact on the organization in various ways. Some of the ways in which these are implemented are: by policies and procedures; by administrative actions relating to organization, personnel, operations, crises and emergencies, training, promotions, rewards and punishments, labor relations, and the host of other matters with which management is concerned; and by comments and expressed attitudes. The policies, actions, and attitudes of top management representatives collectively create the organizational climate—for better or for worse as the case may be. This climate, in turn, is infused throughout the organization by a process that is somewhat akin to osmosis.

It will be of interest to us to give specific attention to some of the aspects of organizations over which management has some control, with the particular purpose of studying the relationship between such organizational characteristics and the effectiveness of individuals, groups, and the organization as a whole.

ORGANIZATIONAL STRUCTURES AND RELATIONSHIPS

Most organizations have a formal organizational structure that delineates such matters as the formal responsibilities and lines of authority of individuals and organizational units. Aside from the formal structure, informal groups and relationships frequently evolve. These groups and relationships may develop for the purpose of exercising power in the organization, or for the purpose of getting things done more efficiently. In the latter case, the by-passing of formal relationships and procedures may be an indication of perceived inadequacies of the formal relationships.

One of the factors that may contribute to organizational success is its organizational structure. In general, it probably can be said that the

organizational structure should be one within which human beings can perform most effectively. While certain principles of organizational structure have evolved through the experience of industry, some questions have been raised about the adequacy of some such principles.[15] In this connection, Leavitt [16] perceives organizations as having the following characteristics: pyramidal shape (which tends to increase interpersonal competitiveness), a hierarchical distribution of authority (with its tendency to increase dependency); the idea of individual responsibility, with the assumption that a large and complicated task can be cut down into non-overlapping, individual-sized pieces; and large size, with consequent difficulties of communication. He expresses the conviction that each of these characteristics carries a potential for intensifying conflict and frustration in individuals and for increasing psychological pressures on the manager.

In turn, he suggests certain organizational changes that can help reduce some of the psychological problems that organizations create. At least a couple of these related to organizational structure. Such changes include: the relatively recent trend in industry of decentralization (which shortens and simplifies human communications, and transfers many decision-making functions to lower levels); separation of staff from line (to distribute responsibility more broadly although authority is not so distributed); use of committees; opening up horizontal communications; and providing for some type of internal "sense organ" for detecting organizational conflicts and trouble spots as they develop (a few companies have a "department of organization" that serves such a function)

On the basis of empirical evidence, it is premature to think in terms of firm guidelines or principles relating to organizational structures that would, as Leavitt says,[17] resolve some of the psychological problems that organizations create. It might be pertinent, however, to review briefly two or three aspects that have been subjected to research, not because they provide confirmed solutions to organizational problems, but rather to illustrate some of the current directions of organizational research that might have future applicability.

Group Network Let us take group structure in problem solving as an example. Given various individuals who form a group, the group can be organized in various ways. Some of these are illustrated in Fig. 13.1. In each case the communication relationships are represented by the lines. Certain configurations (as the wheel, "Y," and chain) are centralized in that all communications are forced through a central position; one

---

[15] R. L. Katz, "Toward a More Effective Enterprise," *Harvard Business Review* (September–October, 1960).

[16] H. J. Leavitt, *Managerial Psychology* (Revised edition), (Chicago: University of Chicago Press, 1964), Chapter 24.

[17] *Ibid.*, p. 239.

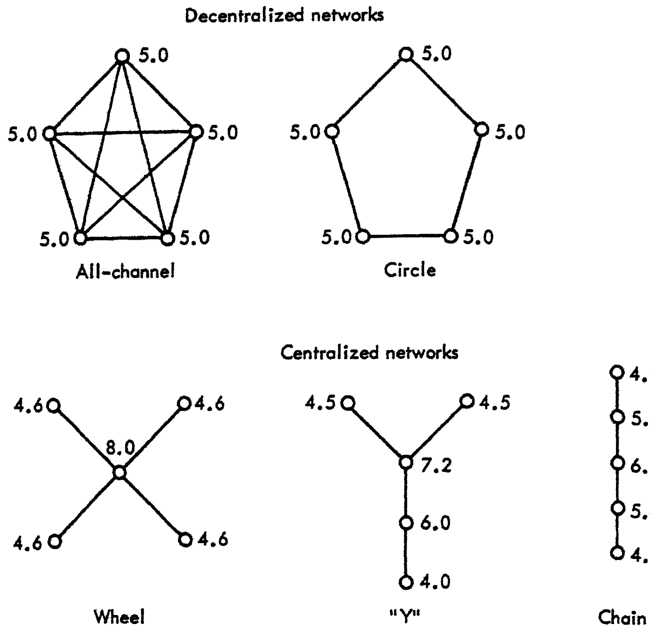


Fig. 13.1. Examples of various types of group communication networks. The lines represent the possible communication channels in a problem-solving task. The numerical values are indexes of "centrality."

member has more channels and more information than the other members of the group. Other networks (as the all-channel and circle) are decentralized in that there is no central member, and all members have equal numbers of channels and the opportunity to share equal amounts of information. An index of "centrality" was proposed by Bavelas; [18] while the formula will not be given, this index is the ratio of the sum of the minimal "distances" of all positions to all others over the sum of the minimal distance of the position in question. Centrality indexes are given for the various positions shown in Fig. 13.1.

In a study by Leavitt [19] the circle, wheel, "Y," and chain nets were used in a laboratory problem-solving task that involved the transmission of written messages among the participants (the nature of the problem need not be described). He found that the circle net (the more de-

[18] A. Bavelas, "Communication Patterns in Task-oriented Groups," *Journal of the Acoustical Society of America*, 22 (1950), 725–730.

[19] H. L. Leavitt, "Some Effects of Certain Communication Patterns on Group Performance," *Journal of Abnormal and Social Psychology*, 46 (1951), 38–50.

centralized) resulted in more errors, required more time, and required more messages than did the others (the more centralized nets). But this is not the complete story; the subjects indicated greater "satisfaction" with the circle net. Subsequent confirmation of this pattern has been reported by Leavitt;[20] and a study by Vermillion[21] revealed greater efficiency of centralized nets (a wheel and a chain) as contrasted with decentralized nets (all-channel and circle) with problems of varying difficulty. As indicated by Costello and Zalkind,[22] centralized nets generally facilitate efficient performance of routine problem solving (especially in the assembling of information), strengthen the leadership position, and result in quickly stabilized sets of interactions among members of the group. On the other hand, decentralized nets rather systematically produce higher levels of satisfaction (probably because they are more "democratic" in nature), and, in addition, seem to facilitate the handling of ambiguous and unpredictable situations, and are more likely to be responsive to innovation.

Examples of various networks as they might exist in industry are given in Fig. 13.2; these might represent either formal or informal groups.

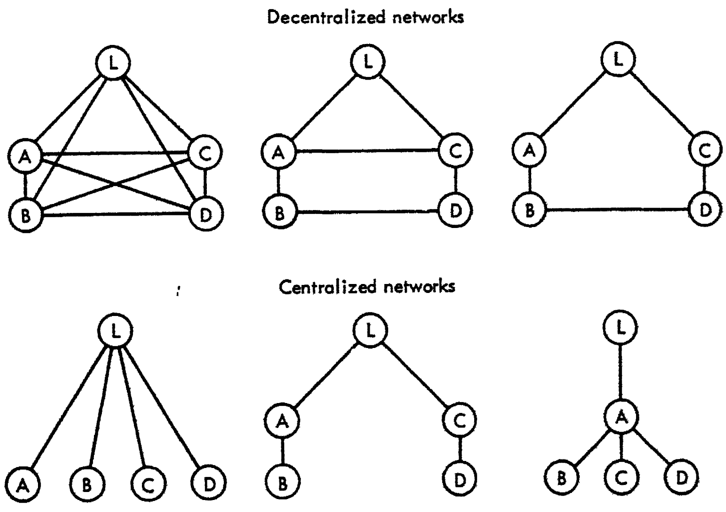


Fig. 13.2. Examples of networks as they might be established in industrial situations The leader is designated as L in each case.

[20] *Ibid.*, p. 232.

[21] W. H. Vermillion, Jr., *Problem Solving Proficiency as a Function of Group Structure and Problem Complexity*, Ph.D. Thesis, Purdue University, January, 1964.

[22] T. W. Costello and S. S. Zalkind, *Psychology in Administration* (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1963), p. 457.

In each case, the leader is indicated by position L. It has been suggested by Leavitt [23] that for small meetings, conferences, and other situations where everyone's ideas are worth something, the "best" networks are those that have two characteristics: they are equalitarian (in that everyone has access to about the same number of channels), and they provide everyone with at least two direct communication channels. The all-channel, circle, and variations thereof generally meet these criteria; these tend to yield higher morale and greater willingness to work. In terms of "efficiency," however, the centralized networks have the advantages of imposing clear-cut organization on the group, defining each person's job, and leaving little leeway for wandering away from that job. It should be pointed out that the data now available relating to group structures come primarily from laboratory studies. As pointed out by Leavitt [24] and as implied in the review by Glanzer and Glaser,[25] one might wonder as to whether such laboratory results have any relevance to situations in the real world of industry. Since very little such research has been carried out in industry, it is, of course, risky to generalize too far. At the same time, laboratory studies such as the above at least suggest the possibility of generalization—subject, of course, to subsequent confirmation from real-life studies

Work-group Linking Pins Another aspect of organizational structure that may have implications in terms of organizational effectiveness is the way in which groups are integrated, including those at various levels within the organizational hierarchy. Earlier reference was made to Likert's *principle of supportive relationships*.[26] This principle implies that an organization will function best when its personnel function as members of highly effective work groups with high performance goals. Starting from this base, Likert [27] proposes that management should deliberately endeavor not only to develop such effective groups, but also to link them into an over-all organization by means of people who hold overlapping group membership; the superior of one group is a subordinate in the next group, and so on. This overlapping relationship is illustrated in Fig. 13.3; Likert refers to it as a "linking pin" function. Partial support for this type of relationship comes from studies by Pelz [28] in which it was found that subordinates expect their superiors to be able to exercise upward influence when dealing with problems which affect the workers

---

[23] H. J. Leavitt, *op cit.*, p. 236.

[24] *Ibid*, p. 235.

[25] M. Glanzer and R. Glaser, "Techniques for the Study of Group Structure and Behavior: II," *Psychological Bulletin*, 58 (1961), 1–27.

[26] R. L. Likert, *New Patterns of Management* (New York· McGraw-Hill Book Company, 1961), Chap. 8.

[27] *Ibid*, p. 105.

[28] D. C. Pelz, "Influence, A Key to Effective Leadership in the First-Line Supervisor," *Personnel* (November, 1952), 3–11.

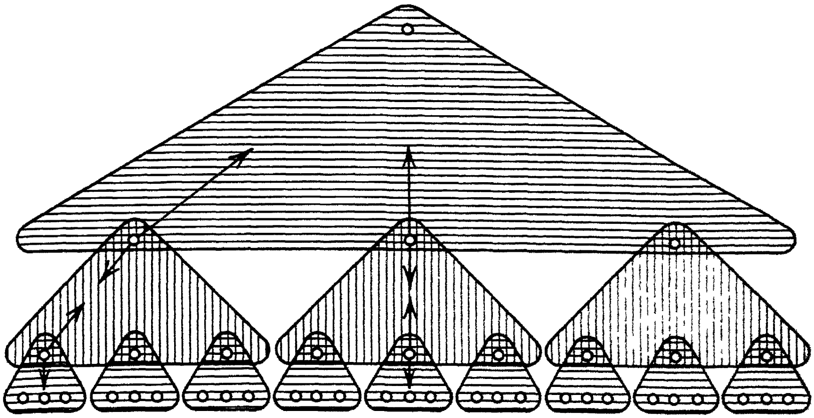


Fig. 13.3. The "linking pin" concept of group organization. The supervisor of one group is also a member of the next higher group, and so on. Communication links are implied within each of the work groups. (Adopted from Likert, *op. cit.*, p. 113.)

themselves; when the supervisor's upward influence is perceived as being limited, the subordinates are likely to have unfavorable reactions.

Needless to say, it is not possible to "legislate" through policy pronouncements that a supervisor will have "upward influence," or that a supervisor will become an effective member of both the group he supervises and the one above. But while these consequences cannot be brought about by management action, it is possible for management to create the kind of situation in which such developments can take place and are encouraged—this, through organization, and policy determinations, but perhaps more through the creation of a climate that encourages such developments.

Functional versus Serial Operations Where there are various steps in an operation, it is possible to organize in either of two ways, as illustrated in Fig. 13.4. In the functional organization, there is an organizational unit that specializes in each function (as A, B, and C), perhaps with two or more groups performing the same function (such as Groups A_1, A_2, and A_3); (this is referred to by Golembiewski[29] as a parallel organization). In the serial organization, one unit is responsible for the entire sequence of operations (in the case illustrated, functions A, B, and C), there being two or more units having the entire sequence. Typical organizational practice tends toward the functional organization, with

[29] R. T. Golembiewski, *Behavior and Organization* (New York: Rand McNally & Co., 1962).

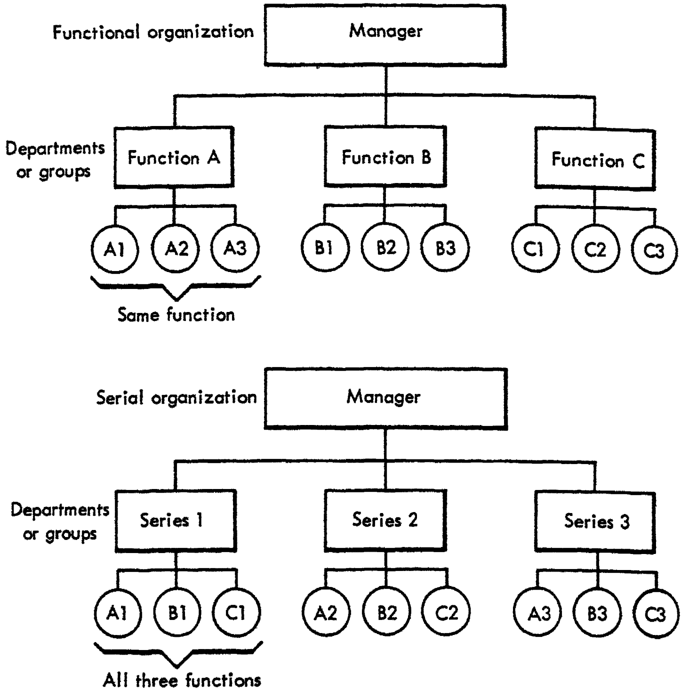


Fig. 13.4. Illustration of two ways of organizing operations with a sequence of steps.

specialization within each unit. It has been suggested by Golembiewski [30] that the serial organization may be more effective in some situations by reason of human motivational factors. In particular, he hypothesizes that with two or more complete units (as in the serial organization) there would be more of a tendency toward feelings of competition between groups, this generating greater group identity, and presumably greater productivity. While some evidence is presented to suggest the implication of greater productivity,[31] it is probable that further research evidence would be required before this hypothesis could be generally accepted.

Discussion As indicated earlier, these examples of organizational research do not now offer panaceas to organizational problems. They may illustrate, however, the current concern in organizational research of the human aspects of organizations, in particular the *organizational* effects of human behavior in different types of organizational structures.

---

[30] R. T. Golembiewski, Paper given on *Psychology in Management Seminar,* Purdue University, October, 1963.

[31] Golembiewski, *op. cit.*

LEADERSHIP

The discussion above has implied that the nature of the leadership in an organization has a very critical bearing on the behavior of individuals in the organization and the effectiveness with which the organizational goals are achieved. It is reasonable to inquire about the validity of this basic assertion.

Leadership and Employee Performance Various studies have indicated, in different ways, the fairly common relationship between the nature of leadership (the leader himself and his leadership "style") and the performance of those being supervised. Most of these studies also bring in other variables, such as the attitudes of the subordinates. One such study is that discussed briefly in Chapter 12, a study carried out by Katz, MacCoby, and Morse[32] in the Prudential Life Insurance Company. It will be recalled that the various sections were characterized as "high" producing or "low" producing sections on the basis of available production records. All of the supervisors of both the "high" and "low" sections were interviewed to determine: (1) a description of supervisory behavior, and (2) attitudes toward own job, subordinates, superiors, the company, and company policies. It was then possible to determine what relationship there was, if any, between supervisory practices and attitudes on the one hand, and section productivity on the other hand.

On the basis of the interviews the supervisors were classified as "employee centered" or "production centered." To help to characterize this distinction, following are the verbatim answers of two supervisors to a question regarding the most important part of the job:

> Supervisor X: "Well, there are two things—keeping the section running smoothly; keeping the clerks happy; keeping production up; making impartial assignments of work; making the proper decisions on some difficult cases involving some payments and maybe a premium that hadn't been paid before." (This supervisor was classified as "employee centered.")
>
> Supervisor Y: "Well, the most important part is to get the reports out. The biggest thing is to get the work out." (This supervisor was classified as "production centered.")

In a similar manner the supervisors were classified as "democratic" *vs.* "authoritarian," and as having "high or average judgment" *vs.* having "poor judgment." The results of comparisons on these classifications for the supervisors of the "high" and "low" sections are shown in Fig. 13.5. This shows that the supervisors of the "high" sections tended to be "employee centered," "democratic," and to have "high or average judgment."

---

32 D. Katz, N. MacCoby, and N. C. Morse, *Productivity, Supervision, and Morale in an Office Situation*, Survey Research Center, Institute for Social Research, University of Michigan, December, 1950.

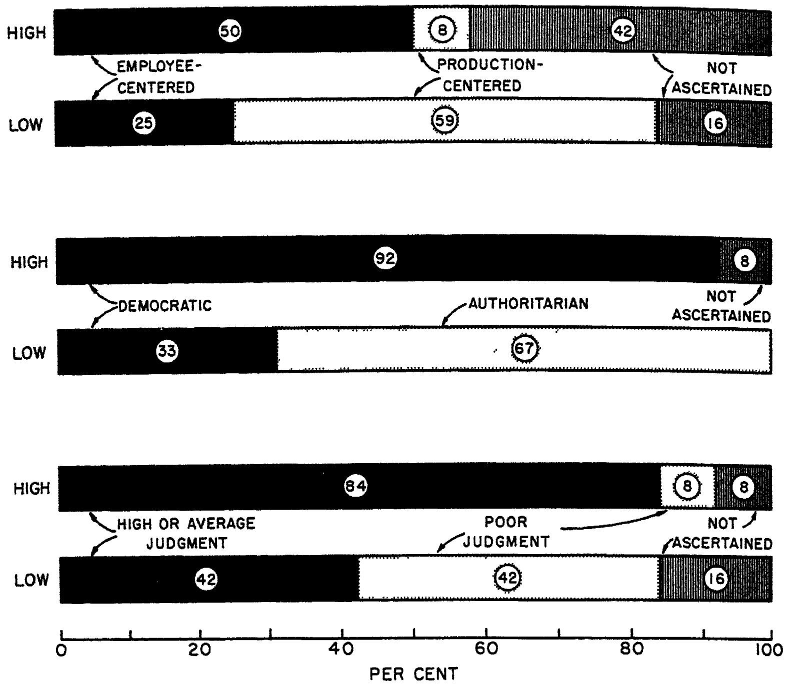


Fig. 13.5. Supervisor characteristics in relation to "high" and "low" producing work sections in insurance company. (Adapted from Katz, MacCoby, and Morse, *op. cit.*)

The supervisors of the "low" sections, in turn, tended more to be "production centered," "authoritarian," and to have "poor judgment." Although it cannot positively be said that these and other associated characteristics were specifically the *cause* of the sections being high or low in production, the weight of logic and inference tend to support this conclusion.

Another study that points to the influences of the supervisor on the performance of the work group is one reported by Lawshe and Nagle.[33] In one plant of the International Harvester Company 208 employees in 14 departments filled out an attitude questionnaire which included 21 items regarding the individual's attitude toward his immediate supervisor. In addition, the 14 departments were rated by six executives on their judgment of work output. This was done by the paired comparison method. The executives were supplied with paired comparison forms and instructed to indicate ". . . the department in each pair which is, in

[33] C. H. Lawshe and B. F. Nagle, "Productivity and Attitude Toward Supervisor," *Journal of Applied Psychology*, 37 (1953), 159–162.

your opinion, doing its job better." Although such ratings do not constitute an objective criterion of department performance, the reliability of judgments among the six executives was very high ($r = .88$), suggesting that there were differences in the performance of the various departments that were perceived quite consistently by the executives.

The results of this study are given in Fig. 13.6. Here we see a rela-

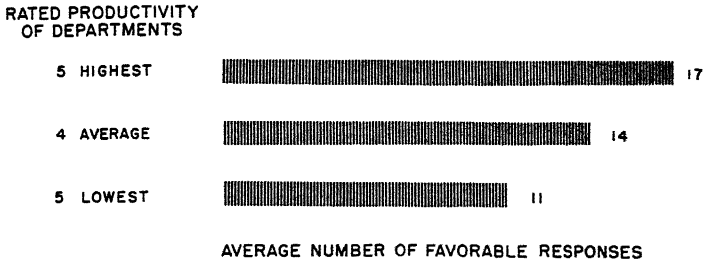


Fig. 13.6. Attitudes toward supervisor (based on responses to attitude questionnaire) of 208 employees in 14 departments of industrial plant in relation to rated productivity of departments. (Adapted from Lawshe and Nagle, *op. cit.*)

tionship between attitudes of employees in the various departments toward their respective supervisors and the rated productivity of the departments. In this study the correlation for the 14 departments between average employee attitude (based on the questionnaire) and rated productivity was .86.

The results of these and other studies indicate quite strongly that the performance of subordinates is influenced to some extent by their leader—whether he is a first-line supervisor, or top executive.

Variations and Dimensions of Leadership It is widely recognized that leadership is not the same thing in every situation. There are obvious variations in the level of leadership positions, in the duties and tasks (even at the same level), in the kinds of jobs under a leader's direct charge, and in the kinds of people supervised. Even if the situation were constant in all such respects, one could find that two or more leaders with entirely different personalities and behaviors could be equally effective, or equally ineffective.

In general, leadership styles have been characterized as falling into three classes: namely, autocratic, laissez-faire, and democratic. As pointed out by Lippitt,[34] the basic difference in these three styles of leadership is in the location of the decision-making function. In the case of the auto-

[34] G. L. Lippitt, "What Do We Know About Leadership?" *National Education Association Journal*, December, 1955.

cratic leadership it resides in the *leader* himself; with laissez-faire leadership it resides in the *individuals*; and with the democratic leadership it resides in the *group*.

Granting considerable situational and individual variation in leadership and leadership styles, one would hope that there would be some common denominators of leadership behavior that have some degree of generality.

Dimensions of leadership behavior. While there are many individual facets of leadership behavior, there appear to be two primary "dimensions" of such behavior. These are described by Fleishman and Harris [35] as follows:

> *Consideration* includes behavior indicating mutual trust, respect, and a certain warmth and rapport between the supervisor and his group This does not mean that this dimension reflects a superficial "pat-on-the-back," "first name calling" kind of human relations behavior. This dimension appears to emphasize a deeper concern for group members' needs and includes such behavior as allowing subordinates more participation in decision-making and encouraging more two-way communication.
> *Structure* includes behavior in which the supervisor organizes and defines group activities and his relation to the group Thus, he defines the role he expects each member to assume, assigns tasks, plans ahead, establishes ways of getting things done, and pushes for production. This dimension seems to emphasize overt attempts to achieve organizational goals.

These two dimensions were developed on the basis of a previous factor analysis of responses to items in the *Supervisory Behavior Description Questionnaire*; the questionnaire is completed by subordinates who "describe" the behavior of their respective superiors. On the basis of responses of subordinates to items on the questionnaire, it is possible to derive separate scores for the supervisors on *consideration* and on *structure*. In general, these dimensions are not negatively correlated (as one might hypothesize); rather, they are relatively independent and uncorrelated. The relative independence of these two dimensions indicates that it is not incompatible for a supervisor to be characterized by any combination of these two dimensions, such as being high on both, low on both, or high on one and low on the other. (This study will be discussed further in the next section.)

The Nature of Effective Leadership Since leadership behavior has been shown to have an influence on the effectiveness of the groups supervised, it is, then, in order to examine differences in such behavior to ascertain if there are certain patterns of behavior that can be identified as contributing to effective performance of those being supervised. Some inklings about this are evident in Fig. 12.1 (discussed above), in which it is evident that the supervisors of high-producing sections (as contrasted

---

35 E. A. Fleishman and E. F. Harris, "Patterns of Leadership Behavior Related to Employee Grievances and Turnover," *Personnel Psychology*, 15 (1962), 43–56.

with low-producing sections) are more employee-centered, are more democratic, and exercise better judgment. A few other studies will be discussed below, these illustrating the use of various criteria of effectiveness.

Degree of supervision. Another aspect of leadership behavior that seems generally to be associated with productivity of subordinates is the degree of supervision—the closeness with which the people are supervised and the freedom they have in such matters as setting their own work pace. In one survey, service workers were asked to indicate the degree of freedom they had for setting their work pace.[36] The 31 departments in which they worked were divided into three groups in terms of the degree of freedom expressed by the workers. Fig. 13.7 shows for

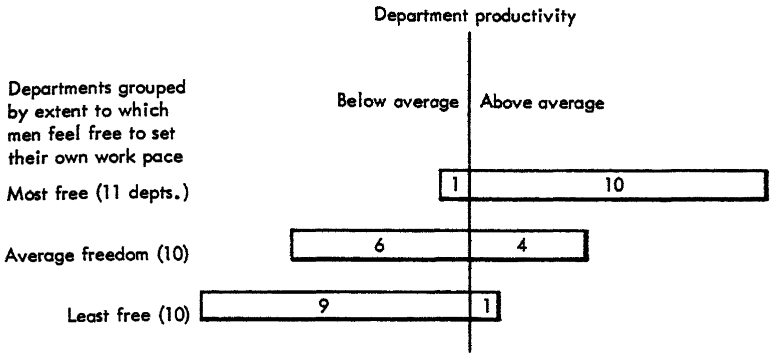


Fig. 13.7. Relationship between freedom service men feel to set own work pace and the productivity of their departments. (From Likert, *op. cit.*, p. 20)

these three groups the number of departments that were above and below average in productivity; this figure indicates a distinct relationship, with the departments in which the men felt "most" free generally being above average in productivity.

As another example, in the study reproduced by Katz, MacCoby, and Morse,[37] the supervisors were characterized in terms of the "closeness of supervision," meaning the degree to which the supervisor checks up on his employees frequently, gives them detailed and frequent instructions, and, in general, limits the employees' freedom to do the work in their own way. (As some employees might put it, "close" supervision is when the supervisor "breathes down your neck all the time.") The analysis of results of the study showed that over 90 per cent of the "low" produc-

[36] R. Likert, *New Patterns of Management* (New York: McGraw-Hill Book Company, 1961), p. 20.
[37] Katz, MacCoby, and Morse, *op. cit.*

tion supervisors gave "close" supervision, whereas only 50 per cent of the "high" production supervisors operated in this way.

In a somewhat corresponding study of railroad workers, Katz, MacCoby, Gurin, and Floor[38] found that the supervisors of the "high" producing railroad gangs (as contrasted with supervisors of the "low" producing gangs): (1) spent more time on supervisory duties than on nonsupervisory duties, (2) were perceived by their employees as generally better in planning; and (3) took more personal interest in their employees. There is also evidence from these and other studies that successful supervisors tend to delegate more authority than those who are less successful.

Consideration and structure in leadership behavior. Referring back to the supervisory dimensions and structure described above, some data are given by Fleishman and Harris[39] from a motor truck manufacturing plant. In particular, consideration and structure scores of the 57 supervisors were related to grievance rates and turnover rates of their respective units. Some of the results are shown in Fig. 13.8. It will be noticed in parts *a* and *b* that, in general, grievance rate and turnover rate increase with *low consideration* scores and with *high structure* scores. The form of the curves is of interest; in each case there is a range within which differences in consideration or structure make no difference. At the extremes (low consideration and high structure), however, there is a marked rise in both criteria.

Parts *c* and *d* of Fig. 13.8 show the interactions. Note especially that *low consideration* is associated with high grievance and turnover rates regardless of the degree of structure. On the other hand, workers under high-consideration foremen (who establish a climate of mutual respect, rapport, and tolerance) are more likely to accept higher levels of structure. Thus, consideration seems to be the more dominant factor.

Leadership behavior as perceived by subordinates. The perceptions by subordinates of the behavior of their superiors do not directly reflect the effectiveness of the work groups in question or the productivity of the individuals. Such perceptions, however, are in accord with other evidence regarding the relationship between supervisory behavior and group productivity, such as illustrated in Fig. 13.6 from the study by Lawshe and Nagle.[40] Aside from this indirect relationship, subordinate attitudes regarding their superiors are significant in their own right.

In a study reported by Katz,[41] 8,000 employees in a large utility company completed an attitude and opinion questionnaire. On the basis of

---

38 D. Katz, N. MacCoby, A. Gurin, and L G. Floor, *Productivity, Supervision and Morale Among Railroad Workers,* Survey Research Center, Institute for Social Research, University of Michigan, 1951.

39 Fleishman and Harris, *op. cit.*

40 Lawshe and Nagle, *op. cit.*

41 D. Katz, "Morale and Motivation in Industry," in W. Dennis (Ed), *Current Trends in Industrial Psychology* (University of Pittsburgh, 1949), pp. 145–171.

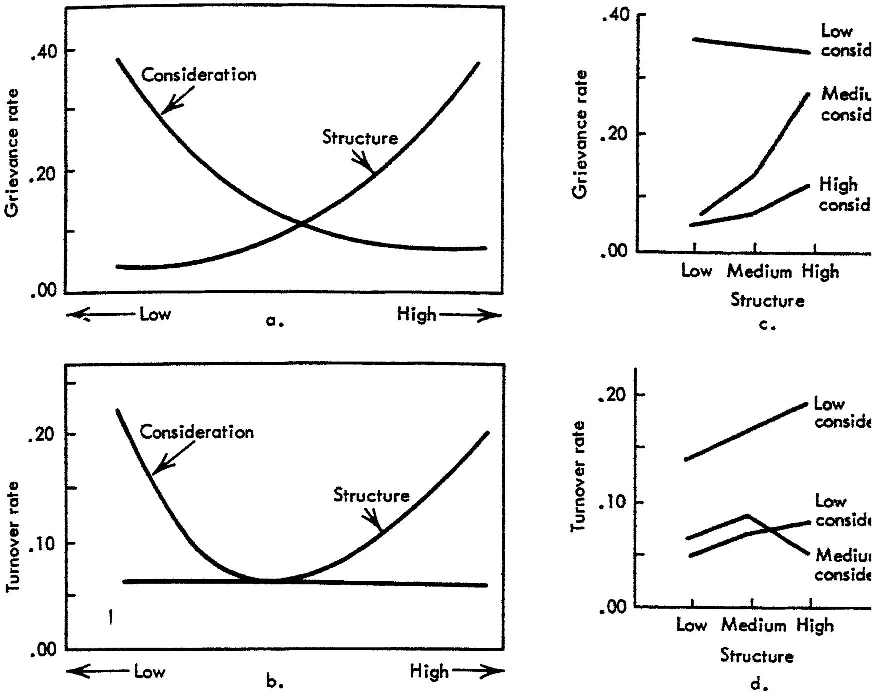


Fig. 13.8. Relationship between degree of *consideration* and *structure* of 57 supervisors and the grievance and turnover rates of their units. (Adapted from Fleishman and Harris, *op. cit.*)

these questionnaires, 40 high-morale groups and 40 low-morale groups were chosen for intensive study. During the interviews, employees were asked, "In what way does your immediate boss supervise you?" In reply they could indicate which of several behaviors were characteristic of their supervisors. For those in the high- and in the low-morale groups, the per cent of employees who characterized their supervisors in terms of each of the various behaviors was determined. Some of these results are given in Fig. 13.9. It will be noted that while there were no appreciable differences in the percentages for the first three behaviors, in the case of all of the others, the percentages for the high-morale groups were higher than for the low-morale groups. These data are generally in accord with evidence from some other studies, and indicate that supervisors of work groups that have generally favorable attitudes tend to take more personal interest in their subordinates, keep them informed, and give them support.

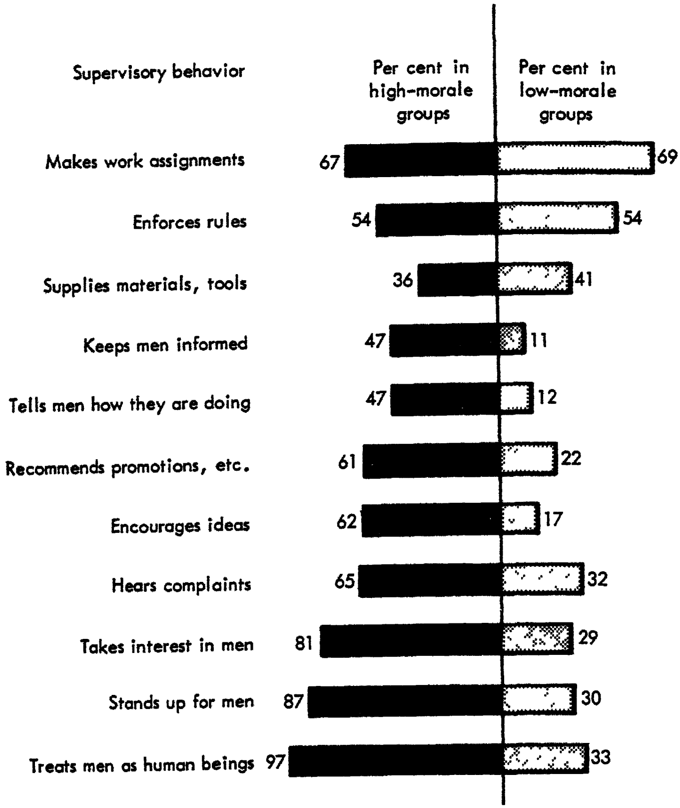


Fig. 13.9. Supervisory behavior as reported by subordinates of 40 high-morale and 40 low-morale groups in a public utility. Each bar represents the per cent of individuals who attributed the behavior to their supervisors. (Adapted from Katz, *op. cit.*, and Likert, *op. cit.*, p. 17.)

A couple of notes of caution are in order in interpreting and evaluating data from such studies. In the first place, there are individual differences in expressed attitudes. In the second place, the "situation" can have a significant relationship to the attitudes of groups. This is indicated by the results of a study by Vroom and Mann [42] in which they found significant differences within the same company in the reaction of employees to authoritarian leadership versus "equalitarian" leadership. Employees in small work groups which were characterized by a great deal of interaction among workers had more positive attitudes toward equali-

[42] V. H. Vroom and F. C. Mann, "Leader Authoritarianism and Employee Attitudes," *Personnel Psychology*, 13 (1960), 125–140.

tarian leadership, as contrasted with employees in large work groups in which there was less interaction; these employees were found to have more positive attitudes toward authoritarian leadership.

Leadership behavior as perceived by their superiors. Subordinates, of course, view the behavior of their superior from their lower rung in the hierarchy of the organization. A view from a rung *above* that of the supervisors in question comes from a study by Roach [43] in which a check list of 328 behavior statements was used. These statements had been gleaned from essay descriptions of 35 "good" supervisors and 35 "poor" supervisors that had been written by their superiors. The completed check list was then used by the managers for 245 white-collar supervisors, and the results were subjected to a factor analysis which revealed groupings of behaviors that tended to occur in combination. The 14 factors that were identified by this process are listed below, in a few cases with brief explanatory comments:

1. *Personal Compliance.* (Degree to which supervisor does what is expected of him, such as setting a good example by being on time, and following through on assignments.)
2. *Job knowledge.*
3. *Direction of group performance.* (Characterized by ability to plan and organize, to carry out procedures, and to see that things function smoothly.)
4. *Rewarding performance and thoroughness of employee evaluation.*
5. *Company loyalty.*
6. *Acceptance of responsibility (decision-making).*
7. *Group spirit.*
8. *Personal drive (motivation).*
9. *Impartiality.*
10. *Poise and bearing.*
11. *Consideration.*
12. *Open-mindedness.*
13. *Cheerfulness.*
14. *Approachability.* (Characterized by extent to which he is willing to sit and talk with his employees, and to which they are willing to talk things over with him.)

Attitudes and Values of Supervisors Still other variables that seem to be associated with effective supervision are the attitudes, self-perceptions, and values of the supervisors themselves. To illustrate this, data from a couple of studies will be mentioned briefly. In one survey, a questionnaire was completed by managers of 31 departments.[44] The questionnaire provided a gross index of the "attitude-toward-men" of the managers, these ranging from unfavorable to favorable. These measures, in turn, were correlated with department productivity indexes, the correlation being

---

[43] D. E. Roach, "Factor Analyses of Rated Supervisory Behavior," *Personnel Psychology*, 9 (1956), 487–498.

[44] The study was conducted by S. S. Seashore, B. Georgopoulus, and A. Tannenbaum, and reported by Likert, *op. cit.*, Chap. 9.

+.64. These results, shown graphically in Fig. 13.10, indicate that the managers of the higher producing departments generally had more favorable "attitudes-toward-men" than the managers of the lower producing departments.

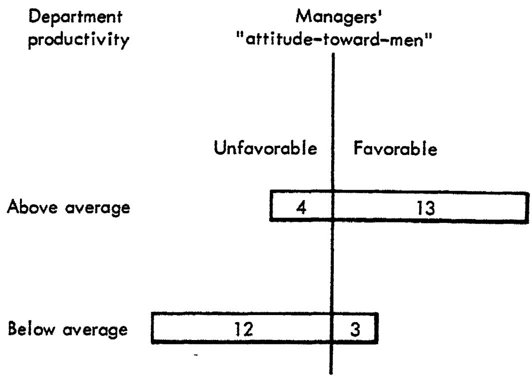


Fig. 13.10. Relationship of "attitude-toward-men" of 32 managers and productivity of their departments. (Adapted from Seashore, Georgopoulus, and Tannanbaum as reported by Likert, *op. cit.*, p. 120.)

Some cues regarding the relative importance of various goals of supervisors comes from the study by Gruenfeld[45] that was reported in Chapter 12. In that study, 52 supervisors rated 18 job characteristics in terms of the importance of the characteristics to themselves; they used the paired-comparison method. The rank order of these characteristics based on the judgments of all supervisors is given in Chapter 12, page 359. From these indications of relative importance of different factors, it appears that the motivation for self-development and actualization looms large in the motivational need hierarchy of the supervisors included in the survey. Motivation for self-development and self-actualization, however, is not synonymous with "looking out" for one's self to the exclusion of other matters. In a report by Dunteman and Bass,[46] for example, it is indicated that supervisors tended to have higher "task orientation" than "self orientation." Task orientation as measured by the *Orientation Inventory* reflects concern about completing a job, solv-

[45] L. W. Gruenfeld, "A Study of the Motivation of Industrial Supervisors," *Personnel Psychology*, 15 (1962), 303–314.

[46] G. Dunteman and B. M Bass, "Supervisory and Engineering Success Associated with Self, Interaction, and Task Orientation Scores," *Personnel Psychology*, 16 (1963), 13–21.

ing problems, working persistently, and doing the best possible job. It might be added that "top" and "middle" performance foremen had higher task orientation scores than "low" performance foremen.

Discussion From these and other analyses of leadership behavior, certain common denominators of effective leadership frequently emerge. This pattern generally is characterized by the following: a positive and genuine interest in, concern for, respect of, and support of, one's subordinates; increased participation by subordinates in the affairs of the group; reasonable freedom of subordinates in certain phases of their jobs; general, rather than close, supervision; more effective two-way communications; and interest in self-development and actualization (but pulling for the company and men rather than for himself).

This pattern, however, is not by any means an entirely consistent one. Individual differences in leadership behavior (even when equally effective) are of major proportions and would be expected. But situational conditions frequently require, or bring about, differences in leadership behavior, these conditions varying with such variables as level and types of personnel involved, the degree of latitude that is available, the organizational functions to be performed, and the nature of the decisions to be made. A case in point is reported by Argyris.[47] The company in question was characterized by high production, low turnover, and favorable employee attitudes. But its employees' attitudes were also characterized by an emphasis on money and security, *non*-involvement in organization problems, and desire *not* to interact with management or employees. In such a situation the organization "mix" and the nature of the leadership involved differ markedly from those of other organizations.

EMPLOYEE PARTICIPATION

In recent years much has been said—pro and con—about the participation of employees in some of the decisions relating to the work situation. Viteles,[48] for example, summarizes the conclusions reported from some of the research on employee participation, as follows: ". . . major outcome of such research is the conclusion that *employee participation in decision-making* in a *democratic atmosphere* created by 'permissive' leadership, facilitates the development of 'internalized' motivation, and serves to raise the levels of the employee production and morale."

---

[47] C. Argyris, "Organizational Leadership," in L. Petrullo and B. M. Bass (Eds.), *Leadership and Interpersonal Behavior* (New York: Holt, Rinehart & Winston, Inc., 1961), pp. 326–354.

[48] M. S. Viteles, *Motivation and Morale in Industry* (New York: W. W. Norton & Company, Inc., 1953), p. 164.

On the other hand, there are those who on various grounds question the desirability of participation. Dale,[49] for example, in reporting the results of a survey among management representatives (as well as union officials and employees) regarding their experiences in formal labor-management cooperation efforts, indicates that the complaints by management about such collaboration were made on the following three grounds: (1) labor was not acting in "good faith"; (2) management functions were challenged or invaded; and (3) the economic costs of cooperation outweighed the gains. It might be added that of the 201 companies with formal cooperation plans that were covered by the survey, the managements of about two-thirds indicated that they were "pleased" with the results of cooperation, with the other one-third indicating varying degrees of neutrality, disappointment, or a "wait-and-see" attitude with respect to the results.

Let us reserve any evaluation of these conflicting views, however, until we have discussed more what we *mean* by employee participation, and until we have discussed certain specific cases.

Types of Employee Participation There are two general types of employee "participation" in industry—namely, what might be considered to be *formal* and *informal.*

Formal participation consists of some plan for labor-management cooperation that is, to some degree, recognized as a *modus operandi* between management and employees, frequently through a union. The degree and character of such "cooperation" varies, as does the field of cooperation. In the survey of 201 plans reported by Dale,[50] for example, there were 22 areas of cooperation. Among the most commonly mentioned areas of such plans were accident prevention, elimination of waste and defective work, furthering labor understanding of policies, attendance, employee insurance plans, quality control, and job evaluation.

The informal type of employee participation occurs more typically at the work-group level, where the supervisor develops the opportunity for the group to share in a problem-solving or decision-making process. Typically, the matters on which such decisions are made are those that are within the range of prerogatives of the leader. In such group decision-making, the leader typically brings the group together, poses the problem to the group, and provides relatively free rein to the group in making a decision. His participation usually is in the capacity of discussion leader, his own input into the decision-making being relatively nominal. It will be recalled that in the discussion of training (Chapter 10) mention was made of sensitivity training. Group decision-making is essentially the same process set in the framework of day-to-day operations as op-

---

[49] E. Dale, *Greater Productivity Through Labor-Management Cooperation,* Research Report No. 14, American Management Association, 1949, p. 26.

[50] *Ibid.*

posed to training. It is a facet of what is referred to as group dynamics.[51]

In characterizing group decision-making, Maier[52] presents the following, indicating both what it *is*, and what it is *not*:

Group Decision Is Not	*Group Decision Is*
1. Abandoning control of the situation.	A way of controlling through leadership rather than force.
2. A disregard of discipline.	A way of group discipline through social pressure.
3. A way of giving each individual what he wants.	A way of being fair to the job and all members of a group.
4. A way of manipulating people.	A way of reconciling conflicting attitudes.
5. A way of selling the supervisor's ideas to a group.	Permitting the group to jell on the idea it thinks will best solve a problem.
6. Sugar-coated autocracy.	A way of letting facts and feelings operate.
7. A matter of collecting votes.	Pooled thinking.
8. Consultative supervision in which mere advice is sought.	Cooperative problem-solving.
9. A way of turning the company over to employees.	A way of giving each person a chance to participate in things that concern him in his work situation.
10. Something anyone can do if he wishes.	A method that requires skill and a respect for other people.

Possible Advantages of Participation The primary theme running through participation in decision-making is that of motivation. In normal circumstances, people become ego-involved in group processes of decision-making into which they have made their input. It is also generally the case that, having participated in making a decision, an individual will also tend to support that decision in his subsequent behavior. Let it be noted, however, that the manner in which the problem is presented, and the way in which the group is "led," can be conducive to such motivation or lead to resentment. Resentment probably would develop, for example, if a supervisor is perceived as manipulating the group to come up with a decision he had already made, or if the group decision is not accepted.

Aside from the possible motivational benefit (and in part based on this), other advantages *may* stem from the use of participation as a managerial procedure. Some of these possible advantages are given by Tannenbaum and Massarik,[53] as follows: (1) a higher rate of output and

---

[51] For an extensive treatment of group dynamics, see D. Cartwright and A. Zander, *Group Dynamics: Research and Theory* (New York: Harper & Row, Publishers, 1962).

[52] N. R. F. Maier, *Principles of Human Relations* (New York: John Wiley & Sons, Inc., 1952), p. 30.

[53] R. Tannenbaum and F. Massarik, "Participation by Subordinates," in R. Tannenbaum, I. R. Weschler, and F. Massarik, *Leadership and Organization* (New York: McGraw-Hill Book Company, 1961), Chap. 7.

increased quality; (2) a reduction in turnover, absenteeism, and tardiness; (3) a reduction in the number of grievances and more peaceful manager-subordinate and manager-union relations; (4) a greater readiness to accept change; (5) a greater ease in the management of subordinates; and (6) the improved quality of managerial decisions. Whether, or to what extent, these possible benefits might accrue presumably would depend upon the interaction of a variety of conditions and situational variables.

Conditions for Effective Participation There is no "formula" for insuring effective participation and its possible benefits; there are, however, certain conditions that are prerequisites, even though they do not insure success of the process. Some of these are set forth by Tannenbaum and Massarik.[54] In the first place, there are certain psychological conditions associated with the individuals. These include the following: (1) the subordinate must be capable of becoming psychologically involved in the participational activities; (2) he must favor the activity; (3) he must see the relevance to his personal life pattern of the thing being considered; and (4) he must be able to express himself to his own satisfaction. The other (non-personal) conditions include the following: (1) time availability (if a decision is urgent, time may not permit group decision-making); (2) rational economics (the decision must be one that is economically sound); (3) subordinate security (his participation must not adversely affect his status or role); (4) manager-subordinate stability (the process must not threaten to undermine the formal authority of the manager, or lead to doubt about the competence of the manager); (5) provision for communication channels through which employees may take part in the process; and (6) education of the participants regarding the function and purpose of the enterprise.

The general pattern described above is reasonably in line with the dimension of consideration that was mentioned earlier. It will be recalled, however, in the study by Fleishman and Harris[55] it was reported that consideration and structure are relatively independent dimensions of leadership behavior.

Developing Effective Leadership The development of effective leadership is, of course, essentially a learning process. In the industrial context, learning can occur as the consequence of organized training, or as a byproduct of everyday work experience. In "learning" to become more effective in leadership functions, it is probable that the most critical change that needs to take place is that relating to attitudes. There is probably no more certain a road to failure in human relations than following *routinely* some set of cookbook "rules" on "how to become a leader in ten easy lessons." Such practices are fairly transparent to those

---

[54] *Ibid.*

[55] Fleishman and Harris, "Patterns of Leadership Behavior," *Personnel Psychology*, 15 (1962), 43–56.

to whom they are applied. What is required is an understanding of others—of why they behave as they do—of their motives and goals. Such understanding—the ability to put one's self in the other person's shoes—must be predicated in large part on attitudinal grounds. In addition, understanding of one's self is important in understanding others.

Training programs for development of leadership ability must, then, provide experiences that *actually* can bring about the modification of attitudes of the participants. Some of the methods mentioned in Chapter 10 are particularly well suited for such training; these include conferences, sensitivity training, case study, and role playing. It should also be recognized, of course, that leadership training frequently covers informational material relating to human behavior—how to deal with others effectively, etc.

Examples of Employee Participation A few examples of employee participation studies will be discussed briefly.

Garment manufacturing company. One of the classic studies of employee participation was carried out at a pajama manufacturing company, the Harwood Manufacturing Corporation, Marion, Virginia, as reported by Marrow [56] and by Coch and French.[57] Because of changes in style trends it became necessary to change the design of the garments being made. A change in garment design, incidentally, necessitates changes in work assignments and the relearning of the new assignments, usually accompanied by reduced piece-rate earnings until the operators get back up to their pre-change levels. Sometimes they never do get back up to their early levels. Because of these factors, there typically is considerable employee resistance to changes in designs.

Since circumstances made job changes necessary, an opportunity existed for an experiment in group participation. In this experiment, four groups of sewing-machine operators were formed, all four groups being matched for the difficulty of the new jobs, for the amount of change in their jobs, and for the level of productivity before the experiment. Three variations in the manner of dealing with these groups were used in the experiment, as follows:

> *Control group.* This group was changed to the new jobs by the normal factory procedure. They were given an explanation of why a change in job methods was necessary, what the new job would be like, and what the new piece rates would be.
>
> *Experimental group 1* A moderate degree of democratic participation was used with this group. They received more information and explanation about the need for change, and they were provided the opportunity to choose representatives who, in turn, participated in designing the new

---

[56] A. Marrow, "Industrial Psychology Pays in This Plant," *Modern Industry,* 16 (No. 1, July 15, 1948), 67ff.

[57] L. Coch and J R. P French, Jr., "Overcoming Resistance to Change," *Human Relations,* 1 (1948), 512–532.

job, setting the new piece rate, and later in training the remaining members of the group. The feeling of participation on the part of the group was such that they spoke of the new job and new piece rates as "our job" and "our piece rates."

Experimental groups 2 and 3. These groups participated directly in designing the job and setting the new piece rates, these having a greater degree of participation than did experimental group 1, which "participated" through chosen representatives.

Records of performance of these groups are shown in Fig. 13.11. The "units per hour" is an index based on standard units of work determined by time and motion study methods. The results, as presented in Fig. 13.11, were quite dramatic. It can be seen for the period of time before the change that all groups were approximately equal. After the

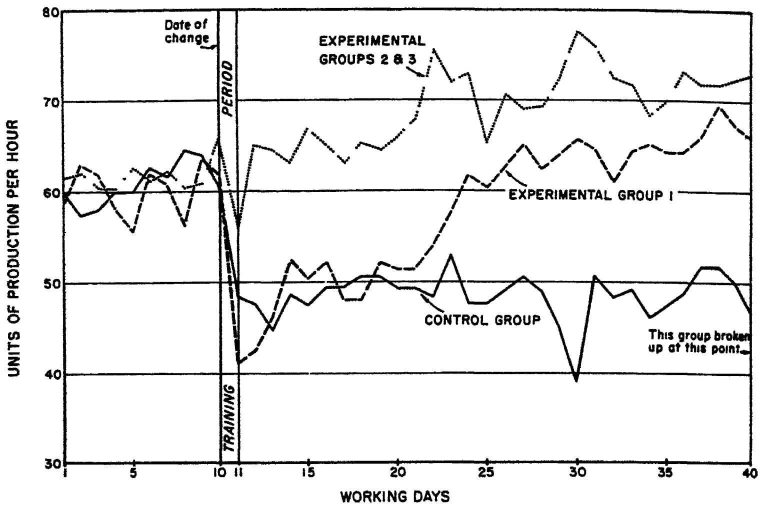


Fig. 13.11. Relationship between degree of participation and work performance of four groups of sewing machine operators. See text for description of participation of groups. (Adapted from Marrow, *op. cit.*)

change the control group dropped to an index of about 50—the usual occurrence after such a change. Many of those in this group quit their jobs. Group 1, which had participation through chosen representatives, dropped to begin with, but consistently recovered over a period of time, ultimately achieving an index of about 65. Groups 2 and 3, which had direct participation, in turn, experienced practically no drop, and continued to increase their performance until they finally levelled off at an index in the low 70's. Not only was the change in productivity after change proportional to the degree of participation of the three groups,

but the turnover rates and amounts of aggression expressed against management were inversely proportional to the degree of participation.

Labor-Management Cooperation Plans. A summary of data for several companies with formal labor-management cooperation plans was prepared by Dale [58] from War Production Board records. This summary included absenteeism rates for the several companies before and after labor participation; the rates are reproduced in Table 13.1. These data show reductions in absenteeism in all nine companies for which data are reported.

Telephone repair crew. In connection with situations in which more informal participation has been used, Maier [59] cites the case of a telephone repair crew consisting of 12 men. The number of repair "visits" varies with the types of repairs to be performed, but these differences tend to average out over a period of time. The average for the crew in question for a six-month period had been between 8 and 9 per day, as compared with the company average of 10.8 visits per day.

Table 13.1

AVERAGE MONTHLY RATES OF ABSENTEEISM BEFORE AND AFTER LABOR-MANAGEMENT COOPERATION PROGRAMS IN NINE COMPANIES

Type of company	*Absenteeism rate (%)* *Before*	*After*
Lamp and stove mfg.	7.0	3.0
Oil refinery	5.3	3.3
Lamp mfg.	6.0	4.0
Ship building	7.0	5.3
Foundry	8.0	4.0
Communications & equipment mfg.	3.1	1.5
Electrical mfg.	10.0	1.0
Engine mfg.	5.0	3.0
Gun and ordnance mfg.	10.0	3.5

Source: *Adapted from Dale,* op. cit., *p. 56.*

In part because of their below-average performance, the foreman held a group meeting to ask his men whether they would care to discuss, as a group, any ideas on how their work could be better coordinated, and any obstacles and difficulties that they thought might be overcome. In doing so, he avoided any criticism of the relative standing of his crew. The group agreed that such a discussion would be worth while.

In their discussions, the crew had various suggestions. One suggestion

58 Dale, *op. cit.*, p. 56.

59 N. R. F. Maier, *Principles of Human Relations, op. cit.*, pp. 225–228.

dealt with *repeats* (repeat repair call). The company had the practice of sending out more highly skilled workmen on the repeats. The group, however, suggested that the man who made the previous repairs be sent back when a repeat call was necessary, and that the foreman go along with the man to help him locate the difficulty. This made it possible for the men to obtain greater experience, and also provided them with the opportunity to know the outcome of their earlier repair work.

Another suggestion consisted of assigning each man to a particular geographical area in the territory, rather than following the previous practice of sending men out on calls over the entire territory. This general scheme was designed to reduce travel time, and to reduce the number of *subsequents* (second or third calls from customers before a repairman arrives).

During the next 2½ years the effectiveness of this plan reflected itself in the records of the work group. Some of these records are summarized in Fig. 13.12. The increase in the number of calls from 8.5 to 12.5 per day, and the reductions in *repeats* and *subsequents* suggest very strongly the effectiveness of group participation in decision-making in this situation.

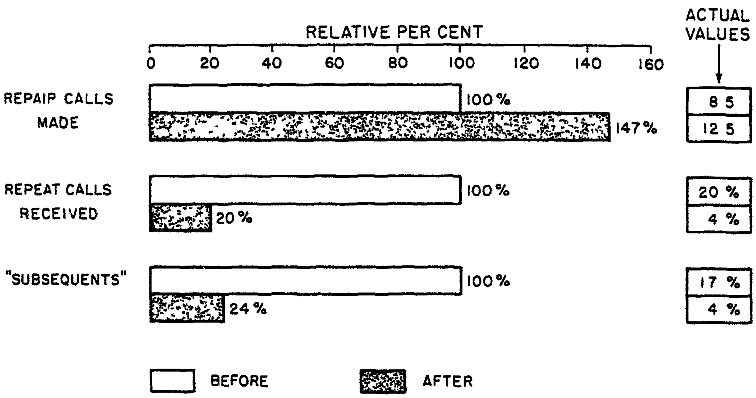


Fig. 13.12. Comparison of performance of crew of 12 telephone repairmen before and after group participation in work procedures. Bars show "before" as 100 per cent, and "after" as per cent of that base. See text for discussion. (Adapted from data by Maier, "Principles of Human Relations," *op. cit*)

Quality of Group Decisions If decisions are to be made, or problems solved, by a group, it is logical to be curious as to the quality of the decisions or solutions. In many situations (such as those discussed above), group decisions or solutions presumably are reasonably satisfactory. The

question of "quality" of decision or solutions is more relevant to problems that require certain kinds of background or experience, or innovation and creativity. A couple of studies of group participation in brainstorming shed some light on this.

In a study carried out by Taylor, Berry, and Block[60] 96 college students participated in problem-solving exercises. Three problems were used, the subjects being asked to come up with as many "solutions" as possible. (In one of the problems, for example, the subjects were asked to suggest what steps could be taken to get more European tourists to come to this country.) Forty-eight of the students were formed into 12 four-man brainstorming groups. The other 48 actually worked individually, but were considered to form 12 "normal" four-man groups, for later comparison purposes. Performance was measured in terms of number of suggested solutions. The mean numbers of unique solutions produced by the real groups and the nominal groups are given below:

	Problems			*Mean of*
Groups	*A*	*B*	*C*	*means*
Real	7 5	17 7	7.3	10 8
Nominal (individuals)	13.6	28.1	17.5	19 8

In this experiment the mean number of solutions pooled from four people *working independently* was nearly twice that of four people *working as a group*. In addition, in an additional phase of the analysis, it was found that the "quality" of the solutions of the real groups was significantly less than of the nominal groups. It was concluded from the study that group participation tends to inhibit creative thinking.

A somewhat comparable study is reported by Dunnette, Campbell, and Jaastad[61] using 48 research personnel and 48 advertising personnel. They used the same problems as Taylor *et al.*,[62] plus one other problem. Half of each group were first formed into four-man brainstorming groups (in which they worked on two of the problems), and then worked individually (on the other two problems); the sequence was reversed with the other half of the subjects in each group. The data for the individuals were "pooled" for groups of four individuals (as with the "nominal" groups of the previous study). The mean numbers of unique ideas or solutions

---

[60] D. W. Taylor, P. C. Berry, and C H Block, "Does Group Participation When Using Brainstorming Facilitate or Inhibit Creative Thinking?" *Administrative Science Quarterly*, 3 (1958), 23–47.

[61] M. D. Dunnette, J. Campbell, and K Jaastad, "The Effect of Group Participation on Brainstorming Effectiveness for Two Industrial Samples," *Journal of Applied Psychology*, 47 (1963), 30–37.

[62] Taylor *et al., op. cit.*

generated by the real groups and by the nominal groups (individuals) are given below:

	Research personnel	*Advertising personnel*	*Combined groups*
Real groups	110.2	97.1	103.6
Nominal groups (individuals)	140.5	141.4	140.9

The quality of solutions also favored the nominal groups.

This study substantially confirmed the one by Taylor *et al.*,[63] namely, that in creative thinking, the pooled individual efforts of many people tend to be more productive of ideas (both in numbers and quality) than group participation.

On the basis of a review of many studies, Lorge, *et al.*[64] indicate that group performance is frequently better than that of the average individual, but is seldom better than the best individual.

Discussion The brainstorming studies mentioned above, and the observations of Lorge *et al.*,[65] have a sobering effect after the more encouraging results of the studies cited earlier. As noted by Wilensky,[66] in summarizing the results of a number of studies of employee participation, "participation" has had both positive, and neutral or negative, effects in workplaces that are both union and nonunion, big and small, prosperous and marginal, and among employees of various social class levels. These comments certainly reinforce the point made earlier that much remains to be learned about the variables that are related to employee behavior.

Granting, however, that much remains to be learned about employee participation, it is obviously not possible to stop the wheels of industry to wait until some further answers are forthcoming from research. In attempting to assess, on presently available evidence, the appropriateness of employee participation as a human relations method, there seems to be more on the positive side of the ledger than on the negative side. As with many other personnel practices, however, participation should be used selectively, considering the purpose in mind. If the purpose is that of generating the best decisions or solutions, some question might be raised about the usefulness of participation processes. On the other hand, as pointed out by Maier,[67] group problem-solving may be preferred to individual problem-solving, even though its superior efficiency cannot be

[63] *Ibid.*

[64] I. Lorge, D. Fox, J. Davitz, and M. Bruner, "A Survey of Studies Contrasting the Quality of Group Performance and Individual Performance," *Psychological Bulletin*, 55 (1958), 337–372.

[65] *Ibid.*

[66] H. Wilensky, "Human Relations in the Workplace: An Appraisal of Some Recent Research," in *Research in Human Relations* (New York: Harper & Row, Publishers, 1957).

[67] N. R. F. Maier, "Fit Decisions to Your Needs," *Nation's Business*, 48 (1960), 48–52.

demonstrated, when acceptance of the solution is important or when morale is a relevant consideration. In a practical situation, however, where the decision perhaps should be made on an individual basis, it may be feasible to arrange for employees' participation in implementing the decision even though they have not participated in the decision-making as such.

WORK GROUPS

In various contexts in the previous discussions references have been made to the implications of work groups in industrial and other organizations. The interaction of people within a group can have a powerful influence on the behavior of the individuals within the group.

Differences in Group Behavior As pointed out by Wilensky [68] and others, the group can influence the behavior of those in the group in various ways. The other side of the coin is reflected by such behavior as the practice of group restriction of work output. Drucker,[69] in discussing output restrictions, expresses the opinion that open restrictions on output provided openly under the provisions of restrictive union rules and "featherbedding" are "only the part of the iceberg that is above the water. Much more important are the invisible, unwritten, informal restrictions decreed by the custom and common law of every plant," as reflected in the tacit setting of production quotas (by the work group) which a worker would be "ill-advised" to exceed.

The more positive effects of work-group unity are revealed by various studies in which it has been demonstrated that there was a relationship between feelings of group cohesiveness and increases in work performance, or other desirable changes in employee behavior, such as reduction of absenteeism. In the study of employees in an insurance company as reported by Katz, MacCoby, and Morse,[70] for example, the employees in the high-producing sections were more strongly "identified" with their groups than were employees in the low-producing sections. There was also evidence from the electric power company investigation reported by Mann and Baumgartel,[71] of a distinct relationship between group unity and absenteeism. Some evidence of this type is shown in Fig. 13.13. In particular this shows the per cent of employees who expressed favorable attitudes toward their work groups, in relationship to the absenteeism averages of their groups. The patterns of relationship

---

[68] Wilensky, *op. cit.*

[69] P. F. Drucker, *The New Society* (New York: Harper & Row, Publishers, Inc., 1950), p. 83.

[70] Katz, MacCoby, and Morse, *op. cit.*

[71] F. Mann and H. Baumgartel, *Absences and Employee Attitudes in an Electric Power Company*, Survey Research Center, University of Michigan, December, 1952.

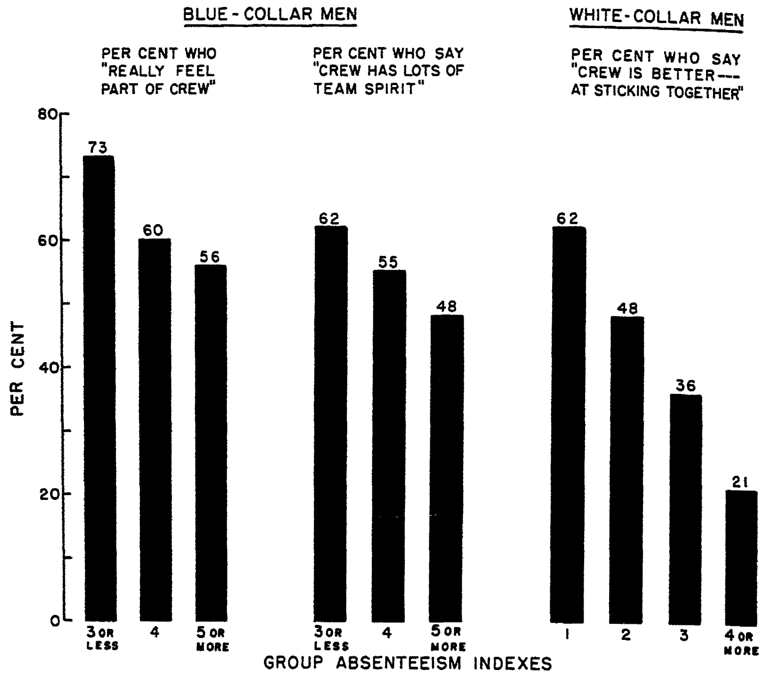


Fig. 13.13. Relationship between absence indexes and attitudes of employees toward their work groups in electric power company. (Adapted from Mann and Baumgartel, *op cit.*)

are somewhat similar for both blue-collar and white-collar men; in work groups with high absence rates there were fewer who expressed feelings of "group unity" than in work groups with low absence rates.

Group Integration The cohesiveness of groups varies greatly from group to group, some being very loose, others being tightly knit. While different definitions of group integration have been proposed, Stogdill [72] defines the real test of integration as the ability to maintain structure and function under stress. The stress presumably can be either internal or external. From the point of view of the individual members, integration is high when the members are loyal to the group, are willing to make strong efforts to support it, and are closely agreed on the goals of the group and methods of attaining the goals.[73] The participation in group activities by individuals is attributed by Stogdill [74] to the "expectancy" of the individuals, defined as a "readiness for reinforcement." He sug-

[72] R. M. Stogdill, *Individual Behavior and Group Achievement* (New York: Oxford University Press, 1959), p. 213.

[73] *Ibid.*, p. 217.

[74] *Ibid.*, Chap. 4.

gests, further, that it is a function of the individual's *drive,* of the *desirability* to him of the possible outcome (the "reinforcement") of his participation in the group, and of the *probability* of that outcome. The nature of the expectancy for individuals obviously varies with one's value system. One person might be active in a group because it bolsters his sagging ego, another because it offers the opportunity to exert authority, another because he wants to work off his aggressions, another because of the social interchange, etc. It would be expected (as Stogdill [75] points out) that individuals tend to seek affiliation with other persons who are perceived to have the same value systems as their own—and who might then "reinforce" their own value systems.

Group Influence upon Individuals Members of a group are subject to pressures from the group to conform to certain standards, patterns of behavior, or (expressed) opinions. This pressure toward conformity has been well documented by various people,[76] and we need not give supporting evidence here. Such pressures apply, however, in varied types of group situations, such as in school, political life, social situations, and work groups. In connection with group pressures, Festinger [77] provides evidence to suggest that the stronger a member's attraction to a group, the greater his tendency to conform to group norms and the greater the pressure that the group can apply to an individual without driving him out of the group. The ways and means of pressuring members of a group to conform to group norms vary in subtlety and degree. These pressures can be positive (to reward conformity) or negative (to punish deviation). A verbal jibe or a cold shoulder can do much to bring a deviating member of a group back into line. It is a durable individual who can withstand the pressures of a group for which he has a strong affinity.

In an industrial organization it is the pressures of the various groups upon their members that bring about behaviors that are compatible with the goals of the organization, or that are at odds with such goals. These differences in group behavior, then, stem basically from the nature of the group goals. Where group goals tend to correspond with those of the organization, the behavior of those in the group will tend to contribute toward the objectives of the organization, and vice versa.

Implications for Management As indicated above, the potential effectiveness of a group within an organization in influencing the behavior of individuals varies with the degree of integration of the group. The possibility of group influence being contrary to the interests of management has tended to cause the managements of some concerns to view dimly the development of strong, integrated groups within the organ-

[75] *Ibid.,* p. 108.

[76] See references such as those listed by Stogdill, *ibid.,* pp. 291–335.

[77] L. Festinger, "A Theory of Social Comparison Processes," *Human Relations,* 7 (1954), 117–140.

ization. Granting such risks, and recognizing that there may be those who would argue with this point of view, there seems to be adequate justification for proposing that one objective of a human relations program is that of developing a situation in which the employees have a strong sense of group unity, and in which they have goals in common with those of the management. In fact, Worthy [78] expresses the opinion that "One of the measures of the effectiveness of an organization is the extent to which the aims of the informal organization correspond with those of the formal." Continuing in this vein he makes the following statement:

> The manager's problem is not that of "bending the informal work group to the purposes of the enterprise." Rather, it is so conducting the enterprise that the relationship between its purposes and those of the primary work group are clearly apparent

To say this is easier than to do it, and there are no cookbook rules that can be followed. The mutual unity-of-purpose of an organization and of the groups within it is typically the end product of many factors, including some of those mentioned above, such as the nature of the leadership, the opportunity for participation, as well as all of the other practices that make up a good personnel program. But perhaps a particularly critical ingredient is the perception by the employees of the attitudes and motivations of management. To repeat what has been said before, the behavior and attitudes expressed by top management tend to establish the climate of the organization.

Formation of congenial work groups. Although much can be done by a supervisor in building a group spirit with an existing work group, there sometimes is the opportunity to form work groups of individuals who indicate that they would like to work together.

A case of this type is reported by Van Zelst [79] in connection with a construction job in which rows of identical houses were being built. The 38 carpenters and 36 bricklayers on the job had been working together for at least five months, and so they were well acquainted with each other's personality and skill.

Sometime after the construction job was started, an experiment was begun, in which the workers were regrouped into groups of two (or, in some cases, four) on the basis of personal choices of work partners. In doing this, each person was asked to list his first, second, and third choices for a partner. All but eight were paired with one of these three, and these eight "isolates" were paired until they could be successfully incorporated into a mutual-choice group.

Objective records were available on labor cost, materials cost, and

---

[78] J. C. Worthy, "Comments on Mr. Wilensky's Chapter," in *Research in Industrial Human Relations, op. cit.*, pp. 52–53.

[79] R. H. Van Zelst, "Sociometrically Selected Work Teams Increase Production," *Personnel Psychology*, 5 (1952), 175–185.

turnover for the 9-month period *before* this experimental period, and for the 11-month experimental period itself. These data are summarized in Table 13.2, showing that the indices of labor cost, materials cost, and turnover all dropped; the reductions were all statistically significant. The implication is that the formation of the congenial work groups had an effect on the motivation of the workers to the extent that the changes given were thereby brought about.

Table 13.2

COMPARATIVE CRITERIA PERFORMANCE DATA OF WORK GROUPS BEFORE AND AFTER FORMATION OF MUTUAL-CHOICE WORK GROUPS

Variable	*Mean index* Before	*Mean index* After
1. Labor cost—per row of units	36.7	32.2
2. Materials cost—per row of units	33.0	31.0
3. Turnover	3.1	.3

Source. *Adapted from Van Zelst, op. cit.*

COMMUNICATIONS

In recent years there has been an increasing recognition of the importance of communications in industrial organizations. It has been pointed out by Likert[80] that the communication function is widely viewed as one of the most important processes of management.

Communication has been defined by Brown[81] as the "process of transmitting ideas or thoughts from one person to another . . . for the purpose of creating understanding in the thinking of the person receiving the communication." This emphasis on *understanding* focuses attention on the *effectiveness* of communications rather than simply on the mechanics of transmitting information. The processes of talking to unwilling ears, or passing out pieces of paper, do not insure that the "message" gets through to the person or group for whom it is intended.

Communication in Management The process of management involves two primary functions. One of these is decision-making and the other is communicating. In discussing the importance of communications, Drucker[82] makes the following statement:

[80] Likert, *op. cit.*, p. 44.

[81] C. A. Brown, "Communication Means Understanding," *Personnel Administration* (January–February, 1958), 12–16.

[82] P. F. Drucker, *The Practice of Management* (New York: Harper & Row, Publishers, 1954), p. 346.

> The manager has a specific tool: information. He does not "handle" people; he motivates, guides, organizes people to do their own work. His tool—his only tool—to do all this is the spoken or written word of the language of numbers. No matter whether the manager's job is engineering, accounting, or selling, his effectiveness depends on his ability to listen and to read, on his ability to speak and to write. He needs skill in getting his thinking across to other people as well as skill in finding out what other people are after

It has been suggested by Bavelas and Barrett [83] that one can, in fact, view an organization as an elaborate system for gathering, evaluating, recombining, and disseminating information. It follows from this that effectiveness of an organization depends very largely upon the effectiveness of its communications—upon the quality and availability of information within the system.

While most of the concern with communications in industry is focused on those that are *internal* to the organization, it should be noted that virtually every organization also has some communications with individuals and groups *outside* the organization. Some of these communication functions include advertising and selling, training and instruction in the use of products, public relations, employee recruiting, and participation in the activities of business, professional, civic, and other groups.

Effects of communications. There are a couple of ways in which communications have an effect (for better or for worse) on the internal functioning of an organization. In the first place, it can bring about changes in attitude As indicated earlier, leadership, participation, group interaction, and related aspects of an organization contribute to attitude change on the part of personnel in an organization or unit; communication is central to all of these. In the second place, communications are, of course, essential to the on-going operations of an enterprise in order to convey information in connection with policies, procedures, decisions, work schedules, difficulties and emergencies, training, supervision, employee relations, advertising, selling, public relations, and the many other matters and processes that are involved in achieving the goals of an organization.

Formal and informal communications. A distinction typically is made between *formal* and *informal* communications. Formal communications are those that are "official"—that are a part of the recognized communication system that is involved in the operation of the enterprise; these communications may be written or oral. Informal communications are those that are "outside" the formal, recognized communication system, such as the conversations between and among workers, and the "grapevine."

---

[83] A. Bavelas and D. Barrett, "An Experimental Approach to Organizational Communication," *Personnel* (March, 1951), 368.

Directions of communication. Communications can flow in various directions. Many (perhaps most) "formal" communications are transmitted *down* the organization from the originating level to or through the lower levels. Other formal communications flow *up* the organization, or *across* the organization (from people on one organizational level to others on about the same level).

Methods of Formal Communication There are, of course, many different methods of both written and oral communication that are available for use by organizations.

Written communications. Following are some of the types of written communications used internally in industry: memoranda and letters (these may be to individuals, to specified groups, or to all personnel); bulletin boards; employee newspapers; house organs; reading racks; handbooks and pamphlets; reports of various types; training and job manuals; posters; employee suggestions; and attitude and other questionnaires. In its relations with the "outside world" (and depending on the purpose of the organization) the organization might prepare such communications as the following: advertising materials; training and instruction materials (for use with the products or services); employee recruiting brochures; and news releases.

Oral communications. Oral communications within the organization include: person-to-person contacts of all kinds (including those concerned with the work of the organization, and those of a conversational nature); interviews (including appraisal interviews); conferences; meetings; training sessions; public-address systems, and telephone. In relations with individuals and groups outside the organization, such oral communications might be used as sales contacts, verbal presentations to individuals and groups, participation in conferences, and speeches.

Effectiveness of Communications The effectiveness of communications must be evaluated in terms of criteria that are relative to the objective(s) in mind. In general, the criteria relate either to the effectiveness in transmitting information as such, or to attitude change on the part of personnel.

Comparison of methods in transmitting information. Various general advantages have been expressed regarding written versus oral communications, some of these being listed below:[84]

Advantages of written communications:

Authority.

Accuracy.

Permanence.

Coverage (can be of adequate length, and can be duplicated and distributed to all individuals who should receive it).

---

[84] S. Habbe, "Communicating with Employees," *Studies in Personnel Policy*, No. 129, National Industrial Conference Board, 1952.

Miscellaneous (retention rate from reading has been claimed to be high; written communications are economical).

Advantages of oral communications:

Personal.

Two-way (ideas can be exchanged, questions asked, and so forth).

Flexible (can be adapted to receiver; can be presented with "right" tone of voice, with smile, gesture, and so forth).

Effective.

Miscellaneous (easy, simple, fast; places responsibility for communication where it belongs—on immediate supervisor).

An interesting study on the effectiveness of various methods presents some factual evidence to go by. This study, carried out by Dahle,[85] involved the "transmission" of information in various departments of an industrial plant by five different methods. Later, tests of information were given to the employees in order to determine how much of it they had learned. The results, given in Table 13.3, show that a combination of oral and written methods was most effective, as indicated by the fact that employees with whom this method had been used had an average test score of 7.70, as contrasted with an average of 6.17 for oral only, and 4.91 for written only. (The written messages were duplicated materials passed out to the employees.) The bulletin board and grapevine methods resulted in average scores of 3.72 and 3.56, respectively.

Table 13.3

AVERAGE SCORES ON INFORMATION TEST OF EMPLOYEES WITH WHOM VARIOUS INFORMATION METHODS HAD BEEN USED

Method of communication	*No. of employees*	*Average score on test* *
Combined oral and written	102	7.70
Oral only	94	6.17
Written only	109	4.91
Bulletin board	115	3.72
Grapevine only	108	3.56

* *All adjacent values in this column differ at the 5% level, except the last two.*
Source: *Dahle,* op. cit.

Communications and attitude change. Some indication of the effects of communications upon attitudes of employees comes from an attitude survey conducted in two plants by the National Industrial Conference Board.[86] In Plant B, special attention had been paid to commu-

[85] T. L. Dahle, "Transmitting Information to Employees: A Study of Five Methods," *Personnel,* 31 (1954), 243–246.

[86] Habbe, *op. cit.*

nications, including the inauguration of regular work-unit meetings in which each foreman met with his employees and discussed matters of common interest. In the other plant (plant A) no such program had been developed. A comparison of the responses of employees in the two plants to a few of the questions provides some revealing contrasts. Results from a few of the questions are given in Table 13.4. A comparison of the per

Table 13.4

RESPONSES OF EMPLOYEES IN TWO PLANTS TO QUESTIONS ON ATTITUDE QUESTIONNAIRE

Question and response	*Per cent giving response* Plant A	Plant B
1. In your own experience, who has been of most help in keeping you up to date on company matters?		
My steward	16	22
My foreman	37 *	47 *
A fellow worker	32	21
One of the top men	15	10
2. When you are asked to do something by your foreman, are you told why?		
Almost never told	24	24
About half the time	38	21
Almost always told	38 *	55 *
3. Can you talk things over with your foreman when you want to?		
Yes, I always can	34 *	56 *
Usually I can	42	27
He is generally too busy	6	4
Usually he doesn't want to be bothered	13	13
I can hardly ever talk with him	5	0
4. Have you been able to get your ideas up to the top men?		
Almost always	21 *	49 *
Sometimes	31	28
Hardly ever	32	13
Not interested in doing so	16	10
5. Do you feel a part of your company?		
I feel I really belong	29 *	62 *
I feel I just work here	42	14
Sometimes I feel one way, sometimes the other	29	24

* *This is the most favorable response. A comparison for the two plants on this response points up the differences between them.*
Source: *Habbe,* op. cit., *pp. 36–38.*

cent giving the favorable responses (especially the most favorable response—marked with an asterisk) for items 1–4 indicates that the employees in plant B felt much more on the same "wave length" with their supervisors than was true of employees in plant A. The responses to item 5 indicate a much greater feeling of "belongingness" on the part of employees in plant B as contrasted with plant A (62 per cent *vs.* 29 per cent).

A generally more favorable attitude was expressed by employees of plant B in answering another question (not shown in the table) about their general feelings toward the company. Forty-five per cent of the employees in plant B said it was "one of the best," whereas in plant A only 20 per cent gave this response.

Distortion in Communications In the transmission process some information can be lost or distorted. This is especially the case where the "information" has to be "interpreted" or evaluated· where it is not of a strictly factual nature. The loss or distortion typically is greater with oral than with written material. Brief mention will be made of a few aspects of such loss.

In the first place, as oral information is transmitted from one person to another, there usually is some loss of detail. In addition, some distortion can occur. This is illustrated by the parlor game in which one person transmits a message by whisper to another, and he to another, etc.; the message that ends up frequently has no relationship to the original message. The same phenomenon can occur in an organization where the information is spread by word of mouth. While the distortion may be entirely unintentional or unrelated to the attitudes or values of individuals, it can be influenced by the receiver's frame of reference. Thus, a change in procedures (designed to eliminate a bottleneck in an operation) may be interpreted as a new policy of "putting the screws" on the employees.

Certain other distorting influences have been discussed by Campbell.[87] Some of these are: loss of information of the "middle" parts of a message (as opposed to the beginning and ending); expectancy (messages that are "expected" are received more reliably than those that are not expected); association with reward and punishment (messages that have reward or punishment implications to the individual usually are assimilated better than "neutral" messages); and relevance to prior input (messages that are relevant to previous messages usually are received more accurately than those that are unrelated to previous messages).

Other Influencing Factors Some other factors that influence the effectiveness of communications have been set forth by Scholz.[88] Some of these are given below:

1. The credibility of the communicator and the motives attributed to him have a profound influence on the reception of his messages.
2. The most successful communications are those which reinforce at least some of the audience's beliefs, those which state conclusions as well as premises, and those which call for action.

---

[87] D. T. Campbell, "Systematic Error on the Part of Human Links in Communication Links," *Information and Control*, 1 (1958), 334–369.

[88] W. Scholz, *Communication in the Business Organization* (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1962).

3. People are interested first in people, then in things, last in ideas. Their attitudes and opinions are strongly influenced by the groups to which they belong or want to belong.
4. It is better to communicate information little by little over a period of time than all at once. Repeating a communication obviously prolongs its influence.
5. Short sentences, familiar words and active verbs help to make communications both interesting and persuasive
6. In changing opinion, oral presentation tends to be more effective than the written word
7. Only rarely is it possible for communication, particularly over the short range, to change deep-seated attitudes or beliefs.
8. Mass communication *alone* is hardly ever an effective agent of change.

The Grapevine The grapevine is the communication system of the informal organization of a company. As pointed out by Davis,[89] it arises from the social relationships of people. Its speed and effectiveness as a communication system—especially in the case of real "newsy" bits of information—are well known. While it lends itself to the transmission of unfounded rumor, and to the distortion of true information, it has been estimated that in typical business situations between 90 and 95 per cent of grapevine information is true.[90]

While the grapevine can have its liabilities in an organization, Davis expresses the opinion that it provides certain benefits such as contributing to the development of group identification and interest in work, and serving as a means of upward communication, especially about how people feel about certain situations. In discussing the grapevine, Davis[91] points out that there are three facts that are recognized: it is here to stay; it is a normal part of any organization; and it offers certain benefits. He therefore suggests that an organization learn to live with it. To do so, he suggests such steps as the following: to become aware of it and take it into account, such as in decision-making; to learn something about it; to "tune in" on it to take advantage of its upward-communication benefits; to work with its leaders; and to "feed" it useful facts—facts that it can aid in transmitting and interpreting.

Needless to say, such an unorthodox approach would be feasible only under conditions of mutual trust and confidence. Further, such a policy should be carried out in such a manner as not to undercut the authority of line managers and supervisors in their official-communication functions. This implies that the types of communications that are "fed" to the grapevine should be the types that subordinates should reasonably

---

[89] K. Davis, "Making Constructive Use of the Office Grapevine," in K. Davis and W. G. Scott (Eds.), *Readings in Human Relations* (New York: McGraw-Hill Book Co., 1959), p 346.
[90] *Ibid.*, p. 350.
[91] *Ibid.*, pp. 351–352.

expect to receive directly from their supervisors. And, in addition, such information also should be made available in some form through official channels.

Readability of Written Communications The preparation of written communications does not insure that they can "communicate" to those for whom they are intended; they must be written in such a manner that they are understandable by the readers. A method for measuring readability of written material has been developed by Flesch.[92] In the measurement of reading ease, Flesch takes into account the following elements: (1) average sentence length in words, and (2) average word length in syllables. These elements are combined to give the reading ease index as follows:

RE (reading ease) $= 206.835 - .846(wl) - 1.015sl$

where wl (word length) is the average word length in syllables of the passage being considered, and

sl (sentence length) is the average sentence length in words of the passage

The index ranges from 0 to 100 for various samples of writing. The "reading ease" scores tend to follow the pattern shown in Table 13.5.

Table 13.5

PATTERN OF "READING EASE" SCORES

"Reading ease" score	*Description of style*	*Typical magazine*	*Syllables per 100 words*	*Average sentence length in words*
0 to 30	Very difficult	Scientific	192 or more	29 or more
30 to 50	Difficult	Academic	167	25
50 to 60	Fairly difficult	Quality	155	21
60 to 70	Standard	Digests	147	17
70 to 80	Fairly easy	Slick-fiction	139	14
80 to 90	Easy	Pulp-fiction	131	11
90 to 100	Very easy	Comics	123 or less	8 or less

Source: *Flesch,* op. cit., *Table 5.*

An example of the reading difficulty of a section in a typical management-union agreement with a simplified version of the same section written in a much more readable manner has been published by Lauer and Paterson.[93] The following passage is reproduced exactly as it appeared in the original contract:

> The Employer shall not discharge any employee without just cause and shall give at least one warning notice of the complaint in writing to the employee and to the Union, except that no warning notice need be

[92] R. Flesch, "A New Readability Yardstick," *Journal of Applied Psychology,* 32 (1948), 221-233.

[93] Jeanne Lauer and D. G. Paterson, "Readability of Union Contracts," *Personnel,* 28 (1951), 36-40.

given to an employee before he is discharged if the cause of such discharge is dishonesty, major violation of company rules that do not conflict with this agreement, or drinking while on duty.

The contract from which this passage was taken has a Flesch Reading Ease score of 28.3—which classifies it as "Very Difficult" and, according to Flesch, requires a college education or its equivalent for adequate understanding.

Here is a simplified version of the foregoing contract provision:

The Employer cannot fire an employee without just cause He has to write the employee and the Union and warn them at least once about the fault. No warning has to be given to an employee before he is fired for these causes:
(*a*) He is dishonest.
(*b*) He commits a major violation of company rules which do not conflict with this agreement.
(*c*) He drinks while on duty

The simplified version of the contract, from which the above passage was taken, has a Reading Ease score of 72.7, which classifies it as "Fairly Easy." According to Flesch, people with only a sixth grade education can readily read and understand this version. What the original passage says in 70 words (all in one sentence), is said more effectively in the simplified version in only 66 words (broken into six short sentences).

Another common type of communication is the union-management agreement. An analysis of the readability of 59 union-management agreements was made by Tiffin and Walsh.[94] Fig. 13.14 shows the distribution

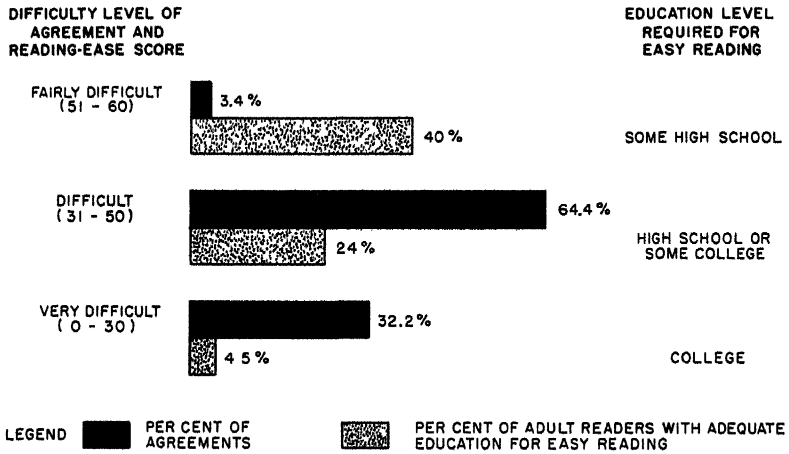


Fig. 13.14. Relationship of reading ease of 59 union-management agreements and per cent of adult readers having educational levels required for easy reading. (Adapted from Tiffin and Walsh, *op. cit.*)

[94] J. Tiffin and F. X. Walsh, "Readability of Union-Management Agreements," *Personnel Psychology*, 4 (1951), 327–337.

of reading-ease scores of these agreements, along with a comparison of the estimated per cent of adult readers having attained the indicated educational level. This indicates very clearly that 96.6 per cent of the agreements were written at levels that required high school graduation or some college education for easy understanding, whereas only about 28.5 per cent of the potential adult readers would be expected to have such education.

In a practical situation these indexes can be used to evaluate the reading ease of written communications. If they are then compared with the reading abilities of those who are to read the communications, it is possible to see whether the communications are within, or above, the reading abilities of the people concerned. There seems to be evidence that many of the written communications in industry *are* above the reading-ability levels of employees, as indicated by Colby and Tiffin,[95] Paterson and Jenkins,[96] and Raney.[97]

INDUSTRIAL CONFLICT

Conflict in some form and degree is part and parcel of virtually every facet of human life, and one would, therefore, not expect the industrial scene to be free of it. By and large, people tend to view conflict as an undesirable component of human life, which it indeed can be. History and experience lead one to the conclusion, however, that conflict can lead to changes that would be regarded as desirable in terms of generally acceptable human values. In industry, for example, conflict has undoubtedly brought about improvement in working conditions for employees. The "desirability" of certain changes, however, may be viewed differently by different people. As expressed by Stagner,[98] one needs to ask, "Bad—from what point of view?" and "Good—from what point of view?" When conflict gets out of hand, however, it tends to become destructive and undesirable for all parties concerned. It has been suggested that an objective of management is to see that conflicts remain on the creative and useful side of an invisible, but enormously important, barrier that divides "good" conflict from the "bad." [99]

Types of Industrial Conflict The very nature of industrial organizations carries with it built-in sources of potential conflicts between different

95 A. N. Colby and J. Tiffin, "The Reading Ability of Industrial Supervisors," *Personnel*, 27 (1950), 156–159.

96 D. G. Paterson and J. J. Jenkins, "Communications Between Management and Workers," *Journal of Applied Psychology*, 32 (1948), 71–80.

97 E. T. Raney, "How Readable Are Your Employee Publications?" *Personnel Psychology*, 2 (1949), 437–459.

98 R. Stagner, *Industrial Conflict* (New York: John Wiley & Sons, Inc., 1956), p. 13.

99 *Conflict Management in Organizations* (Ann Arbor, Michigan Foundation for Research on Human Behavior, 1961).

individuals and groups. Individuals may differ in their opinions about organizational policies or procedures, or may be mutually antagonistic in terms of personality differences. Organizational units that have different functions, such as production, quality control, sales, maintenance, and financial control, may find themselves at odds with each other by reason of their respective functions. Perhaps the dominant source of potential conflict arises from differences between management and employee groups regarding such matters as wages, working conditions, and other related matters.

The Basis for Conflict Whatever its specific nature, conflict arises basically from motivational factors. Stagner [100] expresses this point of view when he states that the phenomena of industrial conflict (and its opposite, industrial cooperation) grow out of the needs of individual human beings. In this connection it is pointed out by Stagner that motivation is distinctly an *individual* phenomenon and that there is *no group mind.* Groups cannot feel, see, desire, fear, or hate any object. Individuals, however, can be influenced by group pressures, as discussed in an earlier section, and their motivation can thus be changed in group situations, these changes generally being in the direction of conformity to the group norms. A group of individuals with reasonably common motivation can, of course, be in conflict with, say, another group.

Somewhat related to the motivational origins of conflict is the role of perception. Human behavior is predicated more on the "perceptions" of people than on the basis of objective "facts." The perceptions of people sometimes tend to be distorted, the distortion generally being in the direction of conformity with one's own frame of reference. Selective perception is related to this phenomenon, and refers to the tendency toward "perceiving" only those aspects of a situation that are compatible with one's frame of reference. Thus, if an employee group perceives or interprets a supervisor's discharge of a fellow employee as an act directed toward breaking up the group, the members of the group would tend to react as though this were the supervisor's motive, whether or not this was the case.

The perception of social phenomena is influenced by various factors, including the following: [101] past experience of the individual; the individual's "expectancy" of what might happen; inner needs (the motivational aspect mentioned above); the possible consequences of some action or event; "field structure" (the organized "pattern" of one's perceptions; perception of individual events or circumstances tend to be "fitted in" to a total pattern).

Group and individual behavior. While group behavior stems basically from the motivation and behavior of individuals, the mere fact of

---

[100] Stagner, *op. cit.*, p. 119.
[101] Stagner, *op. cit.*, Chapter 2.

individuals' being members of a group can bring about collective (group) behavior that would not be characteristic of any single individual. Thus, a mob may engage in violence that the individuals would not otherwise carry out. The anonymity to the outside world that individuals experience within groups seems to offer some promise of not being "identified" individually in group activities.

Labor-Management Conflict It is virtually inevitable that there be some conflict between labor and management, since their objectives are often quite different and, in fact, typically are at odds with the objectives of the opposite party. The cleavage between labor and management in certain circumstances has been very profound. Because of the importance of this source of conflict in our economy, efforts to minimize it would seem desirable from the point of view of the public as well as that of the participants. Toward this end, further understanding of the conflict would be useful.

Understanding through research. For better understanding of the labor-management conflict, it is helpful to have relevant information, some of which can be obtained through research. Information that is relevant to the labor-management conflict could embrace a wide gamut, including levels of job satisfaction, variables associated with job satisfaction, motivations of employees and management, effectiveness of communication, communication processes, attitudes and opinions about various matters, and the perceptions of both parties. It is not feasible to cover such topics extensively, but a couple of examples of relevant studies may be useful.

As an example, let us take a look at the matter of the perceptions of certain groups of people relating to labor-management relations. In one survey,[102] people were asked various questions about such matters as wages, unions, etc. Some results from this survey are given in Fig. 13.15. In particular, this shows the percentage of respondents in certain groups who expressed the opinion that "companies have to be forced" to pay higher wages, and that "most companies would like to break the unions." The results show very marked differences among the groups. Note especially the differences between managers versus union members. Data such as these point up the gulf between the perceptions of groups in conflict; regardless of who is right and who is wrong as far as "facts" are concerned, such differences in perception are indeed real, and need to be recognized and dealt with.

As another example, let us summarize briefly the results of an opinion survey among union members regarding their unions. The survey, reported by Uphoff and Dunnette,[103] covered 1,251 members of 13 unions.

[102] *Public Opinion Index for Industry*, May, 1964, Opinion Research Corporation.
[103] W. H. Uphoff and M. D. Dunnette, "What Union Members Think of Unionism," *Personnel*, (January, 1957), 347–352.

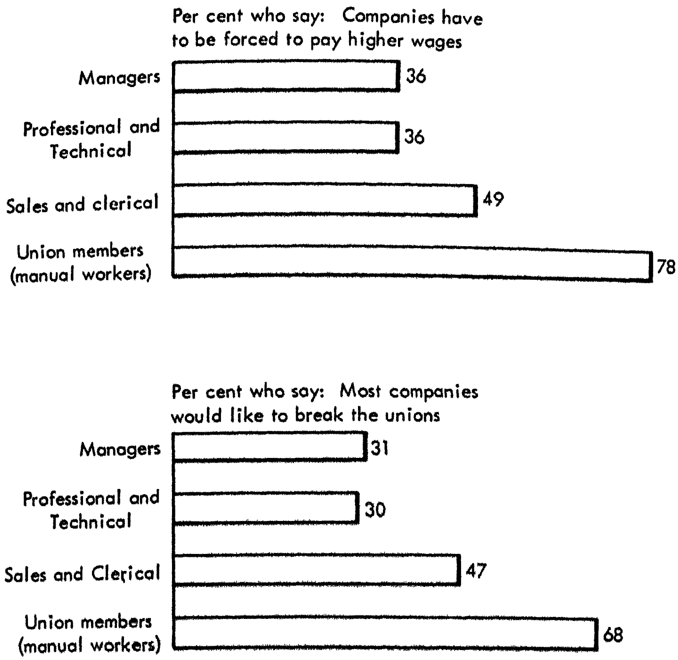


Fig. 13.15. Percentages of individuals in various groups giving specified responses in opinion survey. (Adapted from *Public Opinion Index for Industry, op. cit.*)

The questionnaire used included 77 items. Some of these are given in Table 13.6, along with the percentages of "officers" and "rank and file"

Table 13.6

PER CENT OF UNION OFFICERS AND OF RANK-AND-FILE UNION MEMBERS WHO AGREED WITH SELECTED QUESTIONNAIRE STATEMENTS *

Statement	*Union officers*	*Rank and file*
If it were not for unions, we'd have little protection against favoritism	90	82
Unions should have something to say about the person the employer hires	52	33
Workers should not have to join a union to hold a job	16	38
Fines should be levied for not attending union meetings	57	30
The local union officers are doing a good job	90	73
Our union meetings are run in an efficient manner	77	66

* *Adapted from Uphoff and Dunnette,* op. cit.

who agreed with each statement. Based on these and other data (not given here), the investigators came to the following conclusions: union members agree overwhelmingly that unions protect their rights, and that employees enjoy better wages and working conditions when all of them belong to the union; agreement is far from unanimous on seniority, union participation in hiring, and union-shop provisions; there is need for improved communications within the union; for the most part, the members feel their officers are doing a reasonably good job; many members have no opinions about the national unions; and officers very consistently express more "favorable" opinions than do rank-and-file members.

Aspects of Union-Management Relationships In trying to "understand" union-management relationships it is, of course, useful to be able to "describe" different aspects of the relationship in some manner, in order, for example, to compare one situation with another. To do this, however, requires that one first determine what the different "characteristics" of the relationship actually are. A systematic approach to the identification of the "dimensions" of these relations is revealed by the results of a rather ingenious study by Derber *et al.*[104] While the details of this study need not be described, it was carried out by questionnaire and interview procedures with top management and union officials in 41 establishments in three areas of Illinois. The questionnaires and interviews covered detailed aspects of labor practices, economic status, attitudes of company officials toward the union and union officials (and vice versa), the union contract, contract negotiations, the company and union organizations, economic considerations, and other matters.

A factor analysis of the data revealed "dimensions" of union-management relations. These may be thought of as the attributes that might be appropriate in "describing" the relations between a given union and management. These dimensions are: (1) management satisfaction; (2) local settlement of disputes (establishments high on this dimension report less pressure in negotiations, less legalism, and faster handling of grievances); (3) union satisfaction with relations; (4) union achievement (as in earnings, fringe benefits, etc.): (5) bargaining style (as characterized by such aspects as activity in bargaining, number of proposals made, pressures and threats, etc.); (6) skill of work force (establishments with high skill ratios tend to experience considerable friction in contract negotiations); (7) union satisfaction with achievement; (8) size of establishment; (9) legalism; and (10) effective grievance handling.

While the investigators point out that further experimentation would be necessary to confirm more adequately this approach to the identification of the "dimensions," these are at least suggestive of some of the aspects of union-management relationships.

---

[104] M. Derber, W. E. Chalmers, and R. Stagner, *The Local Union-Management Relationship*, University of Illinois, 1960.

As another phase of the study by Derber *et al.*,[105] the 41 establishments were characterized in terms of each of three variables covered by the questionnaires and interviews, these being:

1. Influence (union influence in union-management relationships).
2. Pressure (extent to which settlements were based on use of pressure, such as work stoppages, slowdowns, movement of the plant, etc.).
3. Attitude (a combination of the "ranking" of management attitude to union and of union attitude to management).

For each of these, all establishments were categorized as "high" or "low" on the variable (with an intermediate "moderate" category in the case of the influence variable). "Clusters" of establishments were then developed, each cluster including those whose patterns of "high," "moderate," and "low" classifications were the same. The more important clusters are shown in Table 13.7. An analysis of the differences among

Table 13.7

PRIMARY CLUSTERS OF 41 ESTABLISHMENTS IN TERMS OF PATTERNS OF UNION-MANAGEMENT RELATIONSHIPS *

	Level of variable			
Cluster **	*Influence*	*Pressure*	*Attitude*	*No of establishments* **
A	High	High	Low	8
B	Low	Low	High	6
C	Moderate	Low	High	6
D	Moderate	Low	Low	5
E	High	Low	High	4

* *Adapted from Derber,* op. cit., *p 61*
** *The clusters with less than four establishments are not included in this table.*

these clusters (based in part on other data) led to the following characterizations:

Cluster A. Aggression and resistance (a situation in which an aggressive union has the upper hand and management is unhappy about it).

Cluster B. Quiescence (a situation in which the union has little influence and, apparently, no desire for more).

Cluster C. Joint participation (a situation somewhat like B, but union influence is somewhat greater).

Cluster D. Repressed hostility (a situation in which union influence is moderate, pressure is largely absent, and attitudes are unfavorable on at least one side).

Cluster E. Extensive joint participation (a situation in which the union

105 *Ibid*

has developed a high degree of influence with management approval, characterized by low pressure as such).

These probing efforts to "describe" the nature of union-management relationships seem to offer some promise of bringing about greater understanding of this important aspect of our economic enterprise.

The Management of Conflict Granting that conflict is part of industrial life, it should be the intent of management to create situations in which it does not get out of hand and, conversely, is directed into useful channels.

Methods of conflict management. In conflict situations the possible responses of a group (or an individual) have been characterized by Shepard [106] as ranging across a gamut from suppression through total and limited war to bargaining and problem-solving. This continuum is depicted in Fig. 13.16. These methods are described briefly below: [107]

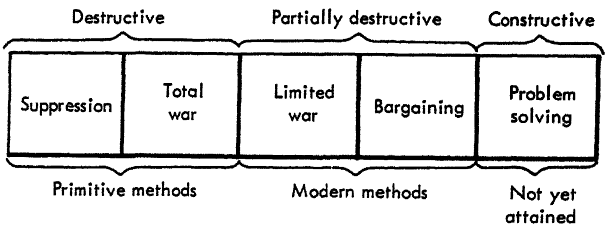


Fig. 13.16. Methods of conflict management, representing a range from destructive to constructive. (Adapted from Shepard, 1962, *op. cit.*, p. 33.)

1. *Suppression.* This method is followed when the two parties are unequal in power, and the stronger uses this to his advantage for the satisfaction of his own interests, using his resources to keep the weaker suppressed. Shepard [108] indicates that an organization chart can be viewed as a suppression chart, in that the pyramid tells us who can suppress whom.
2. *Total war.* This method is normally adopted only when each party believes that it has some chance of winning, but may be used in rare cases when a "liberty-or-death" mentality may lead the weaker party to undertake a foolhardy battle with the strong. Total war is a very primitive form and a very irrational method of conflict resolution, as both sides usually suffer severe losses.

[106] H. A. Shepard, "Responses to Situations of Competition and Conflict," in *Conflict Management in Organizations* (1962), *op. cit*, pp. 33–41.

[107] Adapted from Shepard, *op. cit.*, and from H. A. Shepard, "The Psychologist's Role in Union-Management Relations," *Personnel Psychology*, 14 (1961), 270–279.

[108] Shepard (1961), *op. cit.*, p. 34.

3. *Limited war.* "Limited war" is carried out between conflicting groups under conditions in which the weapons, arena, and methods are all predetermined (sometimes in legal terms), and in which the courtroom may be the arena. The "winner" is determined by some third party such as a judge, jury, or arbitrator. While a decision is made, the underlying conflict usually is not resolved.
4. *Bargaining.* Bargaining lies somewhere between the "limited war" and problem-solving methods. In poor bargaining, both parties feel deprived and each feels that it gave more than it received; the situation may deteriorate to a condition of limited war. In good bargaining each party gives away something that matters little to him and gets something which means a great deal. It can lead to a problem-solving approach to conflict resolution.
5. *Problem-solving.* Problem-solving exists in those encounters in which the parties are able to treat the existence of a conflict as a problem they share in common and have a joint responsibility for solving. It involves a point of view in which the parties regard the advancing of the other's interests as having equal importance with the advancing of its own. In union-management relationships it represents a more advanced state of civilization than we have yet attained, but one toward which we might strive.

Strategies for Dealing with Industrial Conflict While industrial conflict of some type is inevitable, its nature and intensity are to some degree the function of the situation. Since situational variables influence conflict, it is, of course, possible to modify the situation, thereby possibly contributing to the reduction of conflict. While there are no rules-of-thumb that lead to conflict reduction, an organization need not throw up its hands in despair. There are certain positive programs and actions that might reasonably be expected to function in the direction of conflict reduction.

Three basic strategies for dealing with conflict have been set forth by Katz.[109] These are: (1) making the system work; (2) setting up machinery for handling conflict; and (3) changing the organizational structure.

The strategy of trying to make the (present) system work places emphasis on human relations skills in the improvement of interpersonal relations. Some of the matters discussed earlier in this chapter (and in other chapters) are relevant to this effort, as, for example, the development of effective leadership, participation (where appropriate), and effective two-way communications (including feedback to employees). The use of sensitivity training and group dynamics techniques sometimes can bring about some modification of attitudes (and attitudes are, of course,

[109] D. Katz, "Approaches to Managing Conflict," in *Conflict Management in Organizations, op. cit.*, pp. 13–20.

at the core of conflict). The potential effectiveness of all of these is predicated upon the sincerity and integrity of the actions taken by management in carrying them out.

The development of machinery for dealing with conflict places the emphasis on control of conflict rather than on its reduction as such. Such efforts move toward the containment of conflict, thereby reducing the likelihood of spreading. Most typically, provisions are made for resolution of differences at the points where they arise; if they cannot be resolved at that point, they may be carried a step higher, etc., as in grievance procedures.

In some circumstances, changes in the structure of the organization or some segment of it can contribute to conflict reduction by creating conditions in which conflict is less likely to develop. While there are few solid guidelines toward the development of organizational structures that would tend to minimize conflict, Katz [110] offers some suggestions such as: reducing process specialization (to give greater emphasis on the purpose of the organization); decentralization, restructuring to remove obvious differentials in "status" symbols between hierarchical interest groups; development of meaningful cycles of work, with opportunity for employees to "complete" tasks (to avoid tensions from interrupted tasks), and sharing in organizational rewards. Other possible reorganizational possibilities include job enlargement (to increase intrinsic job motivation), reorganization to facilitate communications (both vertical and horizontal), and providing the opportunity for informal groups to develop and exist without fear of possible counteraction.

It might be emphasized here that whatever strategies are adopted should be predicated upon an understanding of human behavior, including the role that motivation and perception play in human behavior in conflict situations.

ADJUSTMENT TO CHANGE

While change is a characteristic of virtually all industrial enterprises, the rate of change in industry has increased markedly as the consequence of the scientific and technological revolution that is going on at a constantly increasing rate. The changes in industry cover a wide range, including: the creation of new products and the obsolescence of others; changes in methods and techniques of production; the creation of new companies, the expansion of some, and the demise of others; changes in methods of marketing and distribution; changes in organizational structures; and others.

---

[110] *Ibid.*

Resistance to Change The frequent tendency in industry (and elsewhere) for people to resist change has been learned the hard way by many individuals and has been confirmed by numerous empirical investigations. Recognizing this, many efforts have been made by individuals and by organizations to deal with such resistance, some such efforts being effective, and others not.

Manifestations of resistance to change. In industrial situations, resistance to change is reflected in various ways, such as slowdowns in work, personnel resignations, lowered morale, strikes or threats of strikes, subtle ways of "sabotaging" the change, personal hostility, and lack of cooperative effort (These symptoms may also be brought about by other conditions.)

Basis of resistance to change. The prospect of some change in an individual's situation typically brings about some speculation—vague or definitive—as to the *nature* of various possible consequences (outcomes) to him; as to the *probabilities* of the various outcomes; and as to the *values* (plus or minus) of the various possible outcomes. If these tend to add up to the strong likelihood of probable gains to the individual, he usually would adopt a favorable attitude regarding the change. If these add up to the strong likelihood of loss to the individual, he would be inclined toward resisting the change. And, if he cannot really figure the odds very well—if he really doesn't have a very good basis for predicting the impact upon him, so that he is left in a state of considerable *uncertainty* about the possible outcomes—he might tend toward taking a dim view of the whole affair

It is thus easy to see how an individual in a job, faced with the prospect of his job being changed, might focus on such possible outcomes as having to learn a virtually new job, and working with new people and a new supervisor; not knowing how well he might be able to make out, he could readily perceive the prospects unfavorably.

Attitudes toward change. It has been suggested by Lawrence[111] that what employees resist is usually not technical change as such but social change—the change in human relationships that generally accompanies technical change. While we cannot fully agree with the major emphasis he places upon the social aspects (as opposed to the technical aspects), one probably should endeavor to ferret out the relative importance of these two facets in any given circumstance.

The attitudes toward technical change of various groups of people in one plant are given in Table 13.8. These data come from a survey by Scott *et al.*[112] that was carried out in a steel mill in Great Britain that

---

[111] P. R Lawrence, "How to Deal with Resistance to Change," *Harvard Business Review,* 32 (No. 2, May–June, 1954), 49–57.

[112] W. H. Scott *et al., Technical Change and Industrial Relations* (Liverpool, England: Liverpool University Press, 1956).

Table 13.8

ATTITUDES TOWARD TECHNICAL CHANGE OF INDIVIDUALS IN VARIOUS OCCUPATIONAL GROUPS IN A STEEL MILL *

Occupational group	*"The firm should put in the new machinery"* Agree %	Not sure %	Disagree %
Middle management	96	4	0
Supervisory staff	94	2	4
Junior staff	76	18	6
Leading hands (production)	72	8	20
Crewmen (production)	49	18	33
Unskilled (production)	50	14	36
Maintenance personnel	64	10	26

* *Adapted from Scott* et al., op. cit., *Table 29, p. 177.*

was undergoing technological change in its processes. (The questions posed dealt with a "hypothetical" plant, but responses to such "oblique" questions generally are indicative of the respondent's own attitudes.) While there were significant numbers of individuals who did *not* believe the new machinery should be installed, it should be noted, in particular, that there were marked differences by occupational groups; the crewmen and unskilled production workers were least favorably inclined toward the change, and were the ones who typically would be affected *most* by changes such as those postulated.

Some indication of the effect of change on social relationships in work comes from the results of a continuation of the above study, as reported by Banks.[113] This phase of the study covered a group of melters, whose work was affected by the installation of new furnaces. About 65 became "redundant" (meaning that they were surplus on the furnaces, but most of them were transferred to other jobs in the plant); the jobs of those who remained on the new furnace were somewhat modified. The attitudes of both groups regarding the effects of the change on their social relationships (with working companions, and team work) are summarized in Table 13.9. In general, in this study more of the post-change melters expressed the opinion that they were worse off than in the case of the redundant melters; but perhaps the most surprising thing about the results was the fairly high percentage in each group who reported "no change." Data such as these are, of course, after-the-fact of change, and

[113] Olive Banks, *The Attitudes of Steel Workers to Technical Change* (Liverpool, England: Liverpool University Press, 1960).

are not necessarily indicative of what the perceptions of the men would have been *before* the change took place.

In this connection, an interesting comparison can be made of the responses of employees from the study by Scott *et al.*[114] and summarized in Table 13.8, with some follow-up data from Banks [115] on the perceptions

Table 13.9

PERCEPTIONS OF MELTERS OF THE EFFECTS OF CHANGE IN THEIR SOCIAL RELATIONSHIPS *

	Change regarding working companions		*Change regarding team work*	
Perception of change	*Post-change melters*	*Redundant melters*	*Post-change melters*	*Redundant melters*
Better	5	15	5	9
No change	64	74	52	77
Worse	31	11	43	14

* *From Banks,* op. cit., *Table 47, p. 81.*

of "over-all" consequences of actual technological change. Table 13.8 reported the following percentages (by occupational groups) of those who "agreed" with the statement that "The firm should put in the new machinery": leading hands, 72; crewmen, 49; and unskilled, 50. Following are the follow-up data from Banks on the percentages who reported that they were "better off" as the over-all consequences of certain changes: leading hands, 92; crewmen, 74; and unskilled, 64. While some caution is necessary in leaning very heavily on such comparisons, they are at least suggestive of the hypothesis that the actual consequences of change may not be as unhappy as they were originally viewed before the change.

It might be added that Banks [116] concluded from the results of her study of steel mill employees that there was no evidence of widespread hostility to technical change as such, although some cautions are in order in extrapolations from this one situation to others.

Dealing with Resistance to Change In the affairs of industrial and other organizations, circumstances arise in which change is virtually inevitable, or is highly desirable. Realizing that resistance might occur, a wise promulgator of change will take such a possibility into account. Lawrence,[117] however, warns us against the *expectation* that resistance

[114] Scott *et al., op. cit.*

[115] Banks, *op. cit.*, Table 32, p. 59.

[116] Banks, *ibid.*, p. 132.

[117] P. R. Lawrence, "How to Deal with Resistance to Change," *Harvard Business Review*, 32 (No. 2, May–June, 1954).

will develop, since any implication of such an expectation might actually touch off the resistance. Rather, he proposes the following point of view: When resistance *does* appear, it should not be thought of as something to *overcome*. Instead, it can best be thought of as a useful red flag—a signal that something is going wrong. But, let us turn our attention to ways and means of dealing with change. As with other management problems involving personnel, there are no easy step-by-step "procedures" one can follow to "reduce resistance to change." Many of the same matters discussed in earlier contexts are relevant to this topic as well—considerations of leadership, participation, communications, etc. Particular mention will be made, however, of certain aspects of some of these processes, as they relate to resistance to change.

Participation. Participation has been extensively promoted as a method for overcoming resistance to change. The study reported earlier by Coch and French [118] and Marrow [119] is a fairly classic case of effective employee participation in industrial change. Some of the results of this study were given in Fig. 13.1. That figure shows the increased production of the groups that "participated" versus the control group. While participation—when used appropriately in appropriate situations—can be an effective means of reducing resistance to change, it is by no means a panacea. A couple of reservations will be mentioned. In the first place, as a decision-making process it is applicable to circumstances where the decision can adequately be made by the group, but many changes need to be based on technological considerations that need to be considered by professionals—engineers, scientists, managers, economists, etc. And, in the second place, if the personnel involved perceive the participation game as being a "manipulation" of them, it is almost certain to fail. "Participation" in the sense implied cannot be achieved simply by bringing people together and presenting them with "the problem"; rather, it can take place only where the basic climate is conducive to it.

Individual interaction. The way in which people are treated frequently affects their attitudes and behavior. Let us illustrate with a couple of incidents that are reported by Lawrence [120] as having occurred in one plant. Both of these involved the prospect of "change" to the same employee.

Incident 1

> We were observing the work of one of the industrial engineers and a production operator who had been assigned to work with the engineer on assembling and testing an experimental product that the engineer was

[118] Coch and French, *op. cit.*
[119] Marrow, *op. cit.*
[120] Lawrence, *op. cit.*

developing. The engineer and the operator were in almost constant daily contact in their work. It was a common occurrence for the engineer to suggest an idea for some modification in a part of the new product; he would then discuss his idea with the operator and ask her to try out the change to see how it worked. It was also a common occurrence for the operator to get an idea as she assembled parts and to pass this idea on to the engineer, who would then consider it and, on occasion, ask the operator to try out the idea and see if it proved useful

A typical exchange between these two people might run somewhat as follows

ENGINEER: I got to thinking last night about that difficulty we've been having on assembling the *x* part in the last few days. It occurred to me that we might get around that trouble if we washed the part in a cleaning solution just prior to assembling it.

OPERATOR: Well, that sounds to me like it's worth trying

ENGINEER: I'll get you some of the right kind of cleaning solution, and why don't you try doing that with about 50 parts and keep track of what happens.

OPERATOR: Sure, I'll keep track of it and let you know how it works.

One day another engineer approached the same operator. The engineer had had no previous contact with the operator, but had been asked to look at one specific problem on the new product because of his special technical qualifications. He had decided to make a change in one of the parts of the product to eliminate the problem, and he had prepared some of these parts using his new method. Here is what happened:

Incident 2

He walked up to the production operator with the new parts in his hand and indicated to her by a gesture that he wanted her to try assembling some units using his new part. The operator picked up one of the parts and proceeded to assemble it. We noticed that she did not handle the part with her usual care After she had assembled the product, she tested it and it failed to pass inspection. She turned to the new engineer and, with a triumphant air, said, "It doesn't work."

The new engineer indicated that she should try another part. She did so, and again it did not work. She then proceeded to assemble units using all of the new parts that were available. She handled each of them in an unusually rough manner. None of them worked. Again she turned to the engineer and said that the new parts did not work.

The engineer left, and later the operator, with evident satisfaction, commented to the original industrial engineer that the new engineer's idea was just no good.

It is from such differences in human reaction as these that Lawrence draws the distinction between the *technical* aspect of the change and the social aspect—the *social* aspect referring to the way those affected by it think it will alter their established relationships in the organization. While the technical aspects of these two incidents were very comparable, the interactions with the engineers were markedly different. When the first engineer approached the operator, she was being treated with the

respect and consideration to which she was accustomed, based in part on the fact that she was a person with some valuable skills and knowledge and some sense of responsibility about her work. In the second incident, the engineer, by his brusque manner, led her to fear that her usual relationships—her customary way of relating herself to others in the organization—were being changed; and she reacted against the way in which she was being treated.

Table 13.10 summarizes the contrasting patterns of these two incidents. While based on these two incidents as such, they tend to illustrate more generally how individuals react differently to differing modes of treatment.

Attitudes In industrial organizations those who typically are responsible for proposing "changes" are staff personnel (engineers, etc.) and line management. The attitudes of such personnel in proposing changes can then affect the way in which the changes are perceived by those who might be affected—as in the above incidents. But the attitudes of such personnel are amenable to some modification. Through example, coaching, training, and other means, top management can help lower levels

Table 13.10

TWO CONTRASTING PATTERNS OF HUMAN BEHAVIOR IN RESPONSE TO PROPOSED CHANGE *

	Change *Technical aspect*	*Social aspect*	*Results*
Incident 1	Clean part prior to assembly	Sustaining the customary work relationship of operator	1. No resistance 2. Useful technical result 3. Readiness for more change
Incident 2	Use new part in assembly	Threatening the customary work relationship of operator	1 Signs of resistance 2. No useful technical result 3. Lack of readiness for more change

* *From Lawrence,* op. cit., *Exhibit 1.*

of management and staff personnel to develop attitudes that would be more likely to generate favorable reactions on the part of those who might be affected by proposed changes. As suggested by Lawrence,[121] for example, staff personnel can sometimes be led to see that winning acceptance of their ideas through better understanding and handling of

121 Lawrence, *op. cit.*

human beings is just as challenging and rewarding as giving birth to an idea.

Discussion The success with which any organization is operated is very much the consequence of the way in which management provides for the effective application of human talent toward the objectives of the organization. Some of the actions that management can take that may contribute toward this end include the development of satisfactory organizations, and the establishment of certain policies and programs such as those discussed earlier in this chapter and in other chapters. The mechanical adoption of "programs" and policies, the use of "gimmicks," and other efforts to "manipulate" people, however, are virtually doomed to failure. The critical ingredient is the climate within which such efforts are made. This climate must be generated by top management.

14

Financial Incentives and Job Evaluation

The factors that cause people to seek employment in certain kinds of work and with certain organizations are varied. Each individual has certain skills and potentialities that he seeks to "sell," and certain sets of personal values and goals that he seeks to fulfill in his work.

Chapter 12 included a discussion of human motivation and some of the job and work situation variables that are related to job satisfaction. On the other hand, the job and organization have certain features which individuals then evaluate against the backdrop of their own values and goals. If we now consider the fact that there are many sellers (people who seek jobs) and many buyers (organizations that are interested in employing people), we have the ingredients of the labor market. Many people do not find what they would consider an ideal job, and on the other hand many employers are not able to employ what they would consider ideal employees. This give-and-take and compromise, however, is part and parcel of the labor market.

One of the primary ingredients in this bargaining process is that of the financial incentives that are associated with the job. While such incentives cover primarily wages and salaries, there are also various fringe

benefits that enter the picture, such as overtime pay, bonuses, insurance plans, stock-ownership plans, and pension plans.

In our discussion we will use the terms *wages* and *wage rates* in the generic sense, to cover earnings whether paid by the hour (wages) or by the week, month, or year (salaries).

THE INCENTIVE VALUE OF EARNINGS

The intent of employing organizations is to offer wages that will attract to the organization personnel who can perform the jobs that are available. Thus, the proffered wages are intended to serve as an incentive to accept employment. Patton,[1] referring particularly to compensation for executives, states: "The compensation structure . . . needs to attract and motivate men having characteristics necessary to success in a particular industry." This same statement is equally applicable to compensation plans for other types of personnel in that, regardless of the type of position(s) in mind, the compensation needs to attract and motivate personnel who have those characteristics that the industry needs. As expressed by Otis,[2] compensation serves as a means of satisfying some of the basic human needs and thus motivating people to expend energy in many ways.

Earnings, however, are considered by employees or prospective employees not independently, but rather in comparison with other features of the job situation. In this connection, for example, Nealey[3] in a series of three studies had employees compare their preferences among certain specified pay and financial benefit conditions. With one group, for example, the employees were asked to indicate their relative preferences for the first eight job variables below, and another group expressed their preferences for all ten:

1. Pay raise of $190 a year
2. Pension increase of $45 a month
3. Eight days of additional paid vacation
4. Family dental insurance ($50 deductible)
5. Medical care for retired employees
6. Twenty days' paid sick leave per year
7. $8,500 additional life insurance
8. Long-term disability pay

---

[1] Arch Patton, *Men, Money and Motivation* (New York. McGraw-Hill Book Company, 1961), p. 67.

[2] J. L. Otis, *The Relationship of the Relative Importance of Functions to Salary Levels within a Company*, California Institute of Technology, Industrial Relations Section, Bulletin No. 30, 1959.

[3] S. M. Nealey, "Pay and Benefit Preference," *Industrial Relations*, 3 (No. 1, October, 1963), 17–28.

9. Short-term unemployment assistance
10. Long-term unemployment assistance

The preferences were obtained by two methods, namely, the paired comparison method and a "game-board" method. In the latter method, the employee was given a hypothetical increase of $190, and was asked to indicate how he would spend it by "allocating" it across the several options. Some of the results of this study for two groups are given in Fig. 14.1. In addition, certain attitude questions were asked, such as "What do you think of the company?" and ". . . of your boss?" Defined by these questions, the low-attitude group tended to prefer pay to benefits, whereas the high-attitude group was more willing to let its "money" ride with the company in various forms of deferred pay-out (pension, hospital, vacation, etc.). Aside from the specifics, these differences start to point up the interaction between management, practices, wages, and morale, emphasizing the point that wages as such cannot be considered independent of other job variables such as other financial and nonfinancial benefits.

It should be added, however, that the nature of the interplay of various financial and nonfinancial considerations—as they serve as incentives—is as yet not at all clear. It has been suggested by Adams,[4] however, that wage administration be viewed within the framework of cognitive dissonance (discussed in Chapter 12). In particular, he proposes that the concept can be applied to wage administration as follows: every man expects a certain relationship between his *input* (what he puts into his work in terms of effort, skill, etc.) and his *outcome* (what he gets in terms of pay and other forms of satisfaction). The cognitive dissonance hypothesis applied to this context includes the prediction that when the degree of dissonance (the difference between the input and outcome) becomes excessive, the individual will take steps to change the situation in some way. Thus if the outcome of one's work (his earnings, other "rewards," working conditions, job satisfactions, etc.) are not perceived as being within the ballpark of equality of one's input, the person might take such actions as asking for a raise, initiating a grievance, lowering his input in terms of his effort, or quitting. In the case of groups, strikes and other actions might occur.

It would thus seem that wage administration policies should be directed toward achieving a balance between two goals: (1) the establishment of wages at those levels that (in combination with other incentives) would generally be perceived as being reasonable outcomes of employees' inputs, and (2) the establishment of wages at levels that are in line with economic considerations of the organization. This balance sometimes can become a tightrope.

---

[4] J. S. Adams, "Wage Inequities, Productivity and Work Quality," *Industrial Relations*, 3 (No. 1, October, 1963), 9–16.

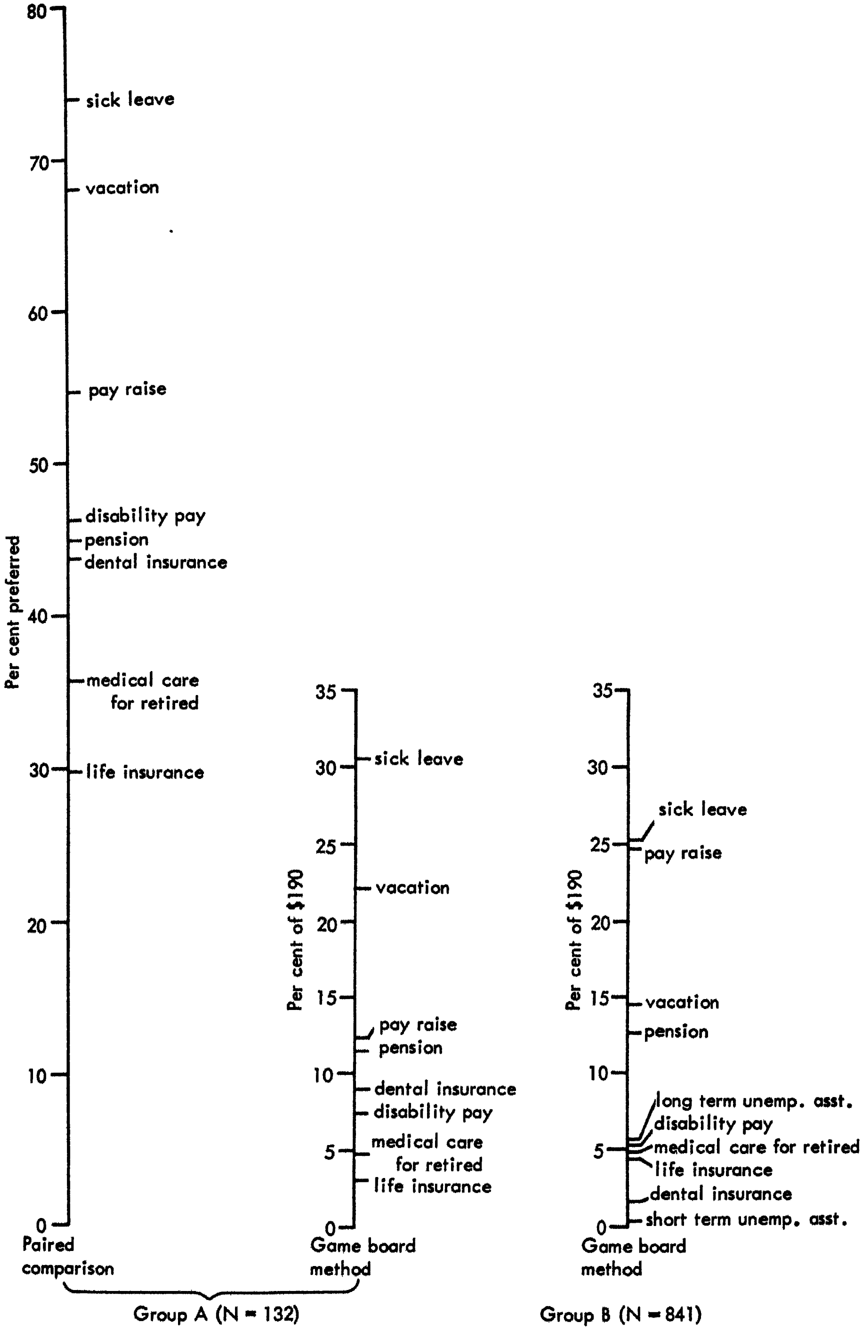


Fig. 14.1. Group preferences for various forms of compensation of two groups of employees. (From Nealey, *op. cit.*)

FACTORS AFFECTING GENERAL WAGE LEVELS

The "general" wage level is brought about by a number of different factors. A few of the more important ones will be mentioned.

Supply and Demand As in the case of other prices, in the long run wages and salaries tend to settle around those levels that are compatible with supply and demand factors. In the absence of any other influences, the effect of supply and demand factors would be that of up-and-down fluctuations that would be related to the relationship between these two factors. In practice, however, other forces such as various laws and other factors have tended to minimize the effects on wages of at least short-time supply and demand forces.

Governmental Wage Controls Various acts of Congress have some bearing upon wages. The most important one is the Fair Labor Standards Act (more commonly known as the Wage-Hour Law). It provides that certain minimum wages be paid by companies engaged in interstate commerce or in the production of goods for interstate commerce.

Contract Negotiations In the case of jobs that are covered by union contracts, rates of pay usually are established during the collective bargaining process. Aside from the influence of this process on the rates for the specific jobs in the specific companies, however, there is some spreading effect to other jobs as well. For example, in the case of some companies that enter into contracts with unions regarding increases in rates for certain jobs, it is sometimes the practice to increase the rates for other jobs by about the same ratio.

Cost of Living Some companies adjust earnings to a cost-of-living index. The most commonly used index for this purpose is the *Consumer Price Index* of the Department of Labor. As pointed out by Lovejoy,[5] however, such an index is not in all respects an adequate basis for determining changes in wage and salary payments. For example, it is more applicable to individuals within certain ranges of standards of living than to others, and is more appropriate for city dwellers than for rural dwellers.

Regional Factors There are noticeable differences in the "going rates" for different regions of the country and different sizes of communities. These may, in part, reflect differences in cost of living in various areas.

Industrial Factors Other factors that influence going rates are those associated with different industries. Some industries tend to pay higher rates for given jobs than do other industries. Such differences may in part be a function of past practice and tradition, of the "newness" of different

---

[5] L. C. Lovejoy, *Wage and Salary Administration* (New York: The Ronald Press Company, 1959), p. 33.

industries, of their relative stability, of competitive conditions, and of differences in the ratios of cost of labor to sale price of commodities. Companies that have high labor costs (relative to sale price) and are in a highly competitive industry may have to watch their pocketbooks much more than those in a low labor-cost industry.

Productivity In the long run, "real" wages (that is, how much people can buy with their earnings) depend very much upon the productivity of industry. It is only by increasing general productivity that the general level of (real) wages can be raised. As pointed out by Lovejoy,[6] the annual increase in productivity from 1919 to 1956 was around 2.8 per cent. This means that during that period the productivity of the country as a whole increased by about 3½ times. This increase is, of course, reflected in lowered relative prices, fewer hours of work, and increased compensation.

SETTING WAGE RATES WITHIN AN ORGANIZATION

An individual organization obtains its employees from some given labor market, or from various labor markets. In general terms, a labor market comprises that usually vaguely defined pool of people who are prospective candidates for employment. For hourly paid and clerical jobs this frequently consists of the community. In some industries the industry itself is the labor market, in that people tend to stay within the industry in question (such as in road construction) and move wherever the employment opportunities seem best. For some management and staff positions, the labor market may be regional or national in scope, but comprising people who have the particular experiences, abilities, or professional qualifications that are required. For example, the labor market for graduating engineers is geographically very broad; many companies send recruiters to engineering schools all over the country. Whatever the labor market in question, a company that expects to obtain its employees from that market must be able to compete with other companies in the market, and wage rates comprise one of the important bargaining ingredients. For a company to be reasonably in line with the labor market in its wage rates, it preferably should approach the process of establishing wage rates in a systematic manner, such as by the use of a job evaluation system.

Job Evaluation Job evaluation is the process of deriving indices of relative job values within an organization on the basis of judgments about the jobs. In turn, the indices of relative job values are used as the basis for determining wage rates of the jobs that are covered by the system. Usually this conversion to wage rates is made on the basis of a wage survey to determine "going rates" of a sample of jobs. Psychologists have an

[6] Lovejoy, *op. cit.*, p. 30.

interest in job evaluation for at least two reasons. In the first place, psychologists are, of course, concerned with human motivation and related matters of the incentives people perceive as fulfilling their needs. And, in the second place, psychologists are interested generally in judgmental processes and in methods of obtaining judgments; in this particular instance, the judgments are, of course, about job characteristics.

Extent of use of job evaluation. Job evaluation systems came into rather common use during World War II, when companies had to be in a position to justify on fairly solid grounds any proposed wage increases. Such systems are far from being universally used, but one survey of executives of 132 companies [7] indicated that about six out of seven of the larger companies surveyed had some kind of job evaluation plan, as did about three out of four of the smaller companies (1,000 employees or less). Of the smaller companies with plans, however, more than two-fifths (43 per cent) had informal plans, as contrasted with only 14 per cent of the larger companies. Some indication of the percentage of companies in which different job groups are covered is given below [8] (the per cents are based only on those companies that have at least one plan):

	Per cent of companies with job evaluation plans	
Employee group	*Larger*	*Smaller*
Plant employees	78	86
Office employees	80	71
Supervisors	51	54
Executives	22	26
Sales employees	28	17

GENERAL PROCEDURES FOR INSTALLING AND OPERATING A JOB EVALUATION PROGRAM

While it is not feasible here to discuss the details that are involved in installing and operating a job evaluation program, some of the major phases of such a program will be described briefly.[9]

Job Evaluation Committee It is usually the practice to establish a job evaluation committee that is assigned responsibility for developing, installing, and administering a job evaluation program. In the survey

[7] *Job Evaluation*, Survey No. 40 of the Bureau of National Affairs, Washington, D.C., December, 1956.

[8] *Ibid.*

[9] The reader interested in further discussion is referred to any of several texts on job evaluation such as:

A. Langsner and H. G. Zollitsch, *Wage and Salary Administration* (Cincinnati: South-Western Publishing Co., 1961).

Lovejoy, *op. cit.*

J. L. Otis and R. H. Leukart, *Job Evaluation*, 2nd ed. (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1954).

mentioned above,[10] for example, about 70 per cent of the companies have such a committee, the remainder placing the responsibility with a single analyst. A committee, of course, can take advantage of the collective talents of its several members, and also provides for participation, which usually is helpful in getting people personally involved in some project.

Job Descriptions for Job Evaluation The first task of the job evaluation committee is to obtain adequate job descriptions. The importance of the latter as a basis for personnel specifications has been emphasized in Chapter 3. Adequate job descriptions are also of vital importance for a system of job evaluation. Jobs cannot be evaluated properly unless their nature is fully known. Just as job titles by themselves are of little or no value in writing personnel specifications, so also job titles must be accompanied by full job descriptions when the jobs in question are to be evaluated.

Evaluation of Jobs Upon completion of the job descriptions, the jobs are then evaluated by whatever system has been developed or adopted by the committee. The various types of systems will be described later. This evaluating process (with most systems) results in the assignment of numerical values, such as points, to individual jobs.

Converting Evaluations to Money Values The next basic process in carrying out a job evaluation program is that of converting the evaluations to money values (hourly wages or salaries). A later section will be devoted to this process, so we will not discuss it in detail here.

Provisions for Evaluation of New Jobs No matter how thorough and satisfactory a job evaluation program is when it is first installed, provision must be made for continual, careful study and evaluation of new jobs as they are created. Jobs are not static because work processes are not static. The job evaluation committee should be a standing committee, and it should meet at regular intervals in order to keep the evaluations current with the creation of new jobs, the changing of old ones, and elimination of some old ones as work processes change.

Getting Jobs in Line after the System Is Installed A problem that is invariably encountered when a job evaluation system is installed for the first time is that some jobs will be found "out of line." These jobs will be of two types—those that have previously had too high a wage rate and those that previously have been set too low.

No problem is encountered in connection with the latter jobs, because the rate increase indicated by the evaluation will, of course, be accepted by the employees on these jobs. It is just as important, however, to fit those jobs that previously have been overpaid into the basic structure. It is obviously unwise, for many reasons, to adopt the policy of cutting these rates abruptly. There are various policies that can be

---

[10] *Job Evaluation*, Survey No. 40, Bureau of National Affairs, *op. cit.*

followed in handling the cases of jobs that are being paid more than the evaluation calls for (these are sometimes called "red circle" rates). On the basis of the survey referred to earlier,[11] the most common method is through normal turnover (deaths, quits, retirements, etc.). About 70 per cent of the companies used this method individually or in combination with other methods. A second method is that of "holding the line" of the out-of-line jobs until the rates of other jobs catch up with them, about half the companies used this method alone or in combination with others. A third method is that of transferring personnel on such jobs to higher rated jobs, and a fourth method is that of changing the job duties so as to upgrade the jobs in question

JOB EVALUATION SYSTEMS

There are four kinds of job evaluation systems that are conventionally described in discussions of job evaluation systems. These are listed below:

Nonquantitative methods
  Ranking method
  Classification method
Quantitative methods
  Point method
  Factor comparison method

The ranking and classification methods provide for the evaluation of jobs in nonquantitative terms, whereas the point method and factor comparison methods result in quantitative job values. These methods and a job description check list method will be described briefly.

Ranking Method The simplest application of this method is through the principle of ranking. Under this plan, all jobs are ranked and placed in a continuum from highest to lowest. The ranking should be done by a person or committee familiar with all of the jobs. In a large plant, where several hundred jobs must be evaluated, this necessary familiarity is seldom possessed by any single person, or even by a group of persons. This is one of the most serious obstacles encountered in the use of the ranking system. It can be minimized to some extent by having each rater rank only those jobs with which he is thoroughly familiar. Then, by a comparison of rankings given jobs common to several raters' lists, it is usually possible to establish key points in devising a scale that will include all jobs.

The reliability of rankings can be increased by using various psychological techniques, the value of which has been proved. Among these are

[11] *Job Evaluation*, Survey No. 40, Bureau of Public Affairs, *op cit.*

the method of paired comparisons (in which the rater compares each job with every other job) and the pooling of rankings of several raters to obtain average rankings.

Classification Method In the operation of this method, a series of job-level classifications or brackets is set up by management or by the committee charged with this function. Each job is then assigned to one of the grades. The system is illustrated by such classifications as junior clerk, senior clerk, principal clerk, and so forth, used in the Civil Service system. In operation, this plan usually results in the grade location of each job being determined largely by the rate that it carried before the job evaluation was made. Langsner and Zollitsch[12] tell of one company that classifies all jobs into one of the following groups: policy-making; administrative; executive; professional; semiexecutive, skilled, and semi-skilled.

While this method is a rather simple one to develop and use, it can permit a tendency to perpetuate any inequalities that may already exist in wage rates, or the evaluations may be influenced by familiarity with present incumbents of the positions in question.

Point Method Point scales are well illustrated by the system used by the National Electrical Manufacturers Association (ordinarily referred to as the NEMA system). The same system is also used by the National Metal Trades Association. This system requires that each job be studied in terms of each of eleven characteristics, and that one of five degrees of each of these characteristics be assigned to the job. The eleven characteristics with the points corresponding to each degree of each are summarized in Table 14.1.

The meaning of each "degree" is clarified by specific "degree" definitions, which are an integral part of this system. For example, in the National Electrical Manufacturers Association system, the "degree" definitions for the various amounts of experience required by the job are as follows:

Degree	*Amount of Experience*	*Points*
1	Up to three months	22
2	Over three months up to one year	44
3	Over one year up to three years	66
4	Over three years up to five years	88
5	Over five years	110

Similar "degree definitions" are included for the various degrees of the remaining ten factors or characteristics. This plan has been specifically constructed for the evaluation of shop jobs. A similar system, encompassing a different list of factors, has been designed for the evaluation of salaried positions; it provides for the rating of salaried jobs on the fol-

[12] Langsner and Zollitsch, *Wage and Salary Administration* (Cincinnati: South-Western Publishing Co., 1961).

Table 14.1

JOB CHARACTERISTICS AND POINT VALUES CORRESPONDING TO VARIOUS DEGREES OF EACH USED IN THE NATIONAL ELECTRICAL MANUFACTURERS ASSOCIATION JOB EVALUATION SYSTEM

Points Assigned to Factors and Key to Grades					
Factors	*First degree*	*Second degree*	*Third degree*	*Fourth degree*	*Fifth degree*
Skill					
1. Education	14	28	42	56	70
2. Experience	22	44	66	88	110
3. Initiative and ingenuity	14	28	42	56	70
Effort					
4. Physical demand	10	20	30	40	50
5. Mental or visual demand	5	10	15	20	25
Responsibility					
6. Equipment or process	5	10	15	20	25
7. Material or product	5	10	15	20	25
8. Safety of others	5	10	15	20	25
9. Work of others	5		15		25
Job Conditions					
10. Working conditions	10	20	30	40	50
11. Unavoidable hazards	5	10	15	20	25

lowing factors: education, experience, complexity of duties, supervision received, errors, contacts with others, confidential data, mental or visual demand, working conditions, character of supervision, and scope of supervision. The last two are used only when supervisory duties are involved, which means that all jobs that do not include supervisory duties automatically receive zero points on these two factors.

Under the operation of this kind of plan, each job receives a total number of points (the sum of the points received for each characteristic). These total point values are then used in setting up the monetary wage scale.

Factor Comparison Method This method has been described in detail by Benge, Burk, and Hay.[13] In this method, fifteen or twenty tentative "key jobs" are first selected. These are jobs that have present rates not subject to controversy and that are considered by the job evaluation committee to be neither underpaid nor overpaid. These jobs are compared with respect to factors common to all jobs. The factors used in the Benge, Burk, and Hay system are:

[13] E. J. Benge, S. L. H. Burk, and E. N. Hay, *Manual of Job Evaluation*, 4th ed. (New York: Harper & Row, Publishers, 1941).

Mental requirements.
Skill requirements.
Physical requirements.
Responsibility.
Working conditions.

The "key jobs" are ranked in order on each of the factors mentioned, and all appear in each of the lists. The rankings are made independently by approximately ten raters, and are made three times by each rater, with periods of approximately one week intervening between each rating.

The going rate (salary or hourly rate) is then divided for each of the tentative key jobs into the amount being "paid" for each of the factors covered by the factor comparison system. This division of the going rate is done without reference to the previous ranking of the tentative key jobs on the several factors.

From these two independent procedures, two rank orders of the tentative key jobs on each factor are determined. The first ranking results from the direct ranking of the jobs on the factors. The second comes from the ranking of the monetary values that have been allocated to each job on each factor following the division of the total rate for each job into the amounts presumably paid for each factor. Any of the tentative key jobs that do not come out with essentially the same rank orders in the two independent ranking procedures are eliminated from the list of key jobs. The jobs remaining constitute the framework of the factor comparison system for the company making the installation. All other jobs in the plant are compared with these key jobs, and each is located in its appropriate place on each of the factors included in the system. The amounts to be paid the job for the various factors are then added, which gives the evaluated rate for the job.

Job Description Check List A method of evaluating clerical jobs reported by Miles[14] involves the use of a check list of activities performed by the job incumbent. It was found in developing this check list that the five operations judged to be most important appeared to be the optimum number to consider for evaluation purposes. For the promotion of good employee relations, all operations performed are included, but the five judged most important are given the most weight.

A definite advantage of this system over the standard point systems in evaluating clerical jobs is that it provides for ease in re-evaluating jobs as they change while a given employee is an incumbent. It is well known that a private secretary may do many things on a job that her predecessor was unable to do. As a new girl develops the ability to handle these "extra" activities, it seems only reasonable that her salary should

14 M. C. Miles, "Studies in Job Evaluation. 9. Validity of a Check List for Evaluating Office Jobs," *Journal of Applied Psychology*, 36 (1952), 97–101.

keep in step with her added responsibilities. The check list job evaluation system if applied to each job at periodic intervals (say every six months) makes it possible to keep salaries in line with actual job operations.

Discussion It should be noted that, of the four conventional systems (excluding check lists), there are differences in the *technique* of evaluation and in the *basis* of evaluation. These differences are illustrated below:

		Technique of evaluation	
		By comparison with other jobs	*By evaluation against a "standard"*
Basis of evaluation	Whole job	Ranking method	Classification method
	Job factors	Factor comparison method	Point method

In terms of frequency of use, the point method is the most common. In the survey previously mentioned,[15] the following percentages were reported (some companies used two or more methods; thus the total adds up to more than 100 per cent):

Method	*Larger companies*	*Smaller companies*
Point method	78%	51%
Factor comparison	29	23
Classification	18	43
Ranking	6	3
Other (including combinations)	6	—

CONVERTING JOB EVALUATION RESULTS TO A WAGE SCALE

Whatever job evaluation system is used, the result is a grading of all jobs according to the principles and assumptions involved in the plan. Ordinarily this grading is entirely completed before any reference is made to the wage or monetary aspect of the work, although in some plans the original evaluations are made in terms of cents rather than points. The experience of most men in this field has shown that original evaluation in terms of dollars or cents is inadvisable, because persons on the job evaluation committee are less likely to evaluate jobs on an impartial basis if they think in terms of money values than if they think in terms of points, the money value of which has not yet been determined.

[15] *Job Evaluation,* Survey No. 40, *op. cit.*

Developing a Going Rate Curve At some stage of developing a wage administration program it is necessary to determine the relationship between evaluations (usually of a sample of key jobs) and rates of pay for corresponding jobs in other organizations in the labor market. A wage or salary survey frequently is carried out in order to obtain such data, if they are not already available To illustrate this, data are given in Table 14.2 for several administrative, technical, and clerical jobs, such as might be found in many companies. This table actually gives median monthly salaries for these jobs, as based on Bureau of Labor Statistics data for the nation (for our purposes, however, we could consider them as being median rates for a sample of companies in the labor market). In addition, and for illustrative purposes, "evaluations" of these jobs are expressed in point values.

To illustrate the conversion of evaluations into rates of pay, the job evaluation points of these jobs are next plotted on Fig. 14.2 in relation to their median salary values; the median is frequently used for this purpose as the "best estimate" of the going rate for each job. The relationship between these two sets of values is then depicted by

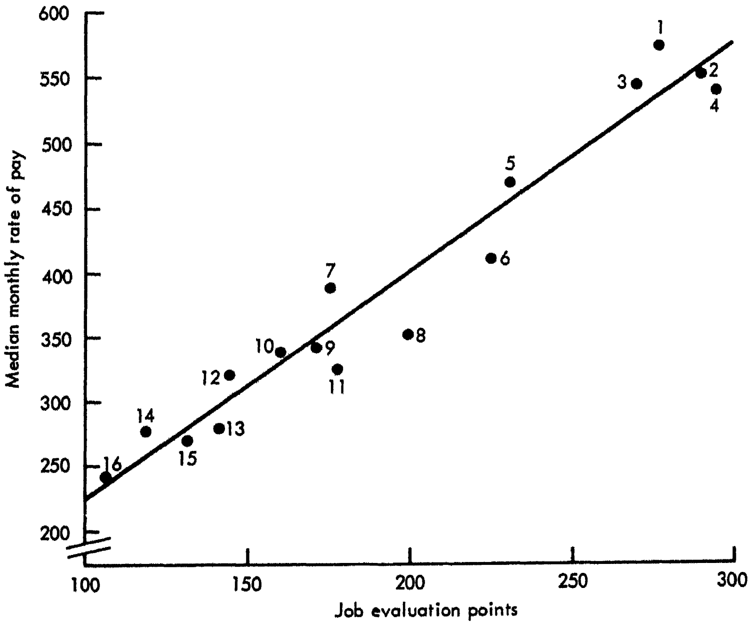


Fig. 14.2. Relationship between job evaluation points and median monthly pay rates for the 16 jobs listed in Table 14.2. The line of best fit (in this case a straight line) is the going rate curve; it reflects the relationship of the job evaluation points to going rates.

a going rate curve; in this particular case the relationship is linear, but with some job evaluation systems the relationship is curvilinear.

Setting Company Wage Curve The next stage is that of establishing a company wage curve. Typically this has some considered relationship to the going rate curve of the labor market; it may be set at the same level, or it may be set at a higher, or lower, level, depending upon such considerations as fringe benefits, contract negotiations, economic considerations, etc. Fig. 14.3 shows a "company" wage curve for the same jobs illustrated in Fig. 14.2; in this case it is shown as being slightly below the going rate curve of the labor market.

Table 14.2

MEDIAN MONTHLY RATES OF PAY FOR SELECTED ADMINISTRATIVE, TECHNICAL, AND CLERICAL JOBS *

Job **	*Job evaluation points*	*Median monthly rate of pay (Dollars)*
1. Job analyst II	277	570
2. Auditor II	290	550
3. Accountant II	269	538
4. Senior draftsman	295	537
5. Tabulating machine operator III	230	466
6. Junior draftsman	225	413
7. Senior stenographer	176	384
8 Bookkeeping machine operator II	198	352
9. Switchboard operator	173	342
10. General stenographer	162	335
11. Typist II	178	328
12. Accounting clerk	145	321
13. Typist I	142	277
14. Bookkeeping machine operator I	119	275
15. File clerk I	133	268
16 File clerk II	107	240

* Wage data from *National Survey of Professional, Administrative, Technical, and Clerical Pay*, Bulletin No. 1346, Bureau of Labor Statistics, U.S Department of Labor, 1962.
** Note: Certain jobs are characterized by Roman numerals These designate different classes of jobs based on job activities and responsibilities, each such class (such as I, II, etc.) is defined separately.

This company wage curve, derived on the basis of the going rate curve, sets the general pattern of rates for all of the jobs covered by the job evaluation system. While this curve is based on an analysis of a sample of jobs that are common to other companies in the labor market, most companies, of course, have some jobs that are unique to them. The use of a company wage curve then insures that all jobs in the company, whether unique to the company or not, will have their rates established on the same basis.

In converting job evaluation points to actual rates of pay, however, there are different practices that are followed. It would be possible to take the evaluated points for a given job, and derive the corresponding exact rate that would be applicable. Thus, every slight difference in points would result in some difference in hourly rate. In practice, most companies (and unions) feel that the inherent lack of perfect accuracy in the judgments that underlie a set of job evaluations makes it desirable to bracket together jobs of approximately the same point value and to consider these jobs as equal in setting up the wage structure. This bracketing results in so-called *labor grades.* The number of labor grades found in specific wage structures varies from around 8 or 10 to 20 or 25. The tendency of most current union demands in wage contract negotiations is to favor a relatively small number of labor grades.

When the jobs have been bracketed in labor grades, provision is usually made for wage increases within each labor grade, as illustrated in Fig. 14.3. Various procedures have been used in granting wage increases within labor grades, as well as in upgrading employees to higher

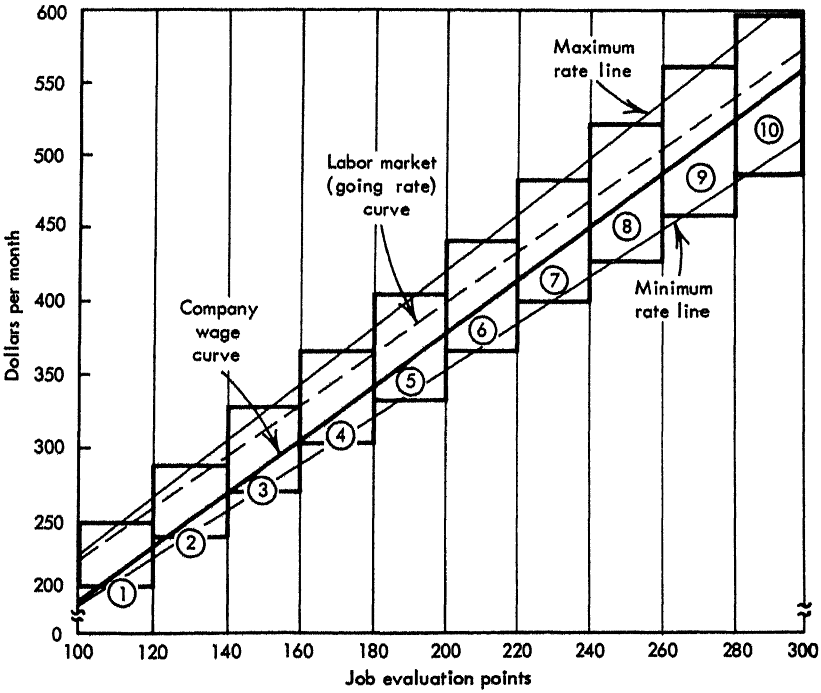


Fig. 14.3. Illustration of a "company" wage curve, and of one pattern of conversion of job evaluation points into company rates. In this particular example, the point values are converted into ten pay grades, each of which has a range of rates.

categories. Some companies use an automatic acceleration schedule under which specified increases automatically become effective after a specified period of time on the job. This principle is employed most frequently in the lower labor grades and with new employees, but it is also sometimes used at higher levels in the wage structure. A systematic merit-rating program is also used by some plants as a means of identifying employees who are eligible for a wage increase under the prevailing wage structure.

The particular method that is followed in making wage increases has often been the subject of controversy in union-management contract negotiations. Unions frequently demand rigid acceptance of seniority as the basic determining factor, and the company as frequently contends that proper managerial responsibility requires the granting of wage increases upon the basis of proved merit rather than their automatic occurrence with tenure on the job. Some companies have used job qualification tests for the upgrading of employees, not only within a given labor grade but from one labor grade to another.

As mentioned above, there are various policies followed in converting points to rates of pay. While these will not be discussed here, two or three variations will be mentioned. Fig. 14.3 shows an increasing range of rates with higher pay grades; in some plans the range may be constant, or may be greater, or less, than that illustrated. In some systems the width of the pay grades (in terms of job evaluation points) increases with higher rates; these increases may be systematic, or adapted in some way to the concentration of jobs along the evaluation scale.

WEIGHTING OF ITEMS IN JOB EVALUATION

Under a point method of job evaluation such as the NEMA plan, different weightings are assigned to the several characteristics. The weights used usually are based upon the experience and judgment of men charged with the responsibility of creating and installing a usable system. In the operation of many plans, different weights are given to the various characteristics by allowing a different maximum number of points for the several items.

In the NEMA system the weights assigned are as summarized in Table 14.3. In actual operation, the functional weightings are determined by the variability of spread of the ratings on each trait. Thus, if the variabilities of the item values were to differ in proportionate value from the maximum values assigned, the actual weightings would not follow the assumed weightings. It is interesting to note, also in Table 14.3, that in three plants the functional weightings computed from variability measures agree rather closely with the weights based on the assigned maximum values. The variable-maximum-value technique

of assigning weights, therefore, would seem to be a fairly effective, as well as a simple method of giving any predetermined set of weights to the factors involved.

Table 14.3

WEIGHTS OF ELEVEN FACTORS ASSUMED BY MAXIMUM VALUES OF ITEMS IN THE NEMA SYSTEM, AND FUNCTIONAL OPERATION OF ACTUAL WEIGHTINGS IN THREE PLANTS

	Item	*Percentage Weights Assumed by Maximum Values of NEMA System*	*Percentage Weights Operating in System in Three Plants* Plant A	Plant B	Plant C
Skill	Education	14	18	17	13
	Experience	22	26	27	26
	Initiative and Ingenuity	14	16	15	14
Effort	Physical Demand	10	6	7	8
	Mental or Visual Demand	5	4	3	5
Responsibility	Equipment or Processes	5	5	5	4
	Material or Product	5	5	3	4
	Safety of Others	5	5	4	6
	Work of Others	5	4	7	4
Job Conditions	Working Conditions	10	8	9	8
	Unavoidable Hazards	5	3	3	8

STATISTICAL STUDIES OF JOB EVALUATION PLANS

Any one of the methods of job evaluation results in a reasonably consistent set of evaluations of the various jobs studied. Each is based upon certain assumptions, which differ from plan to plan, no one system should be considered to be an "exact science." As long as *any* systematic plan is followed that provides for all jobs to be studied in terms of the *same* assumptions and the *same* other stipulations (such as weights for the factors involved), the result is far more satisfactory than a wage structure that has grown piecemeal, without integration. One study[16] has shown that different job evaluation systems give essentially the same results. Chesler found intercorrelations among 6 different company systems to range from .89 to .93. Two of the systems were factor comparison, two were 15-factor point systems, one a 13-factor point system, and

[16] D. J. Chesler, "Reliability and Comparability of Different Job Evaluation Systems," *Journal of Applied Psychology*, 32 (1948), 465–475.

one a combination ranking and classification system. Using from one to three raters in each plant—raters who were familiar with their respective systems—35 standard jobs were rated with the 6 different systems. These results show that essentially similar job evaluation results are obtained with different job evaluation systems.

The choice of which system to use is not, therefore, a problem of choosing from among several plans, only one of which can be correct, all the others being wrong. There are practical advantages as well as disadvantages to every system, and any one plan that is installed with care is almost certain to be better than no plan at all. The time and expense involved in installing and operating a job evaluation system, however, are such that it is advantageous to use a method that will accomplish the desired result with a minimum of labor, if the results obtained from a simple system are equivalent to those that would result from a more complex one.

Certain basic methods of psychology are ideally adapted to a quantitative study of job evaluation systems. These methods have been applied in a series of very significant studies reported by Lawshe and his collaborators.[17] Since the NEMA point system (or a system basically the same) has been more frequently used than any other, Lawshe first made certain statistical analyses of the results of several NEMA-point-system installations to determine whether approximately the same results could have been obtained by the use of a simplified system involving fewer characteristics to be rated.

Using a standard statistical technique,[18] a determination was made of which single factor, which two factors, which three factors, and so on, give point ratings that correlated highest with the total point ratings. The results obtained in three different plants using the NEMA system (or a slight modification of it) are sumarized in Table 14.4.

---

[17] C. H. Lawshe, Jr., and G. A. Satter, "Studies in Job Evaluation: 1. Factor Analyses of Point Ratings for Hourly-Paid Jobs in Three Industrial Plants," *Journal of Applied Psychology*, **28** (1944), 189–198.

C. H. Lawshe, Jr, "Studies in Job Evaluation: 2. The Adequacy of Abbreviated Point Ratings for Hourly-Paid Jobs in Three Industrial Plants," *Journal of Applied Psychology*, **29** (1945), 177–184.

C. H. Lawshe, Jr., and A. A. Maleski, "Studies in Job Evaluation: 3. An Analysis of Point Ratings for Salary-Paid Jobs in an Industrial Plant," *Journal of Applied Psychology*, **30** (1946), 117–128.

C. H. Lawshe, Jr., and Salvatore L. Alessi, "Studies in Job Evaluation: 4. Analysis of Another Point Rating Scale for Hourly-Paid Jobs and the Adequacy of an Abbreviated Scale," *Journal of Applied Psychology*, **30** (1946), 310–319.

C. H. Lawshe, Jr., and R. F. Wilson, "Studies in Job Evaluation: 5. An Analysis of the Factor Comparison System as It Functions in a Paper Mill," *Journal of Applied Psychology*, **30** (1946), 426–434.

C. H. Lawshe, Jr., and R. F. Wilson, "Studies in Job Evaluation: 6. The Reliability of Two Point Rating Systems," *Journal of Applied Psychology*, **31** (1947), 355–365.

[18] The Wherry-Doolittle Technique, one of the basic statistical methods used for this purpose, is described by Stead, Shartle *et al.*, *Occupational Counseling Techniques* (New York: American Book Company, 1940), 245ff.

Table 14.4

CORRELATIONS OBTAINED BETWEEN RATINGS ON SELECTED FACTORS AND TOTAL POINT RATINGS IN THREE PLANTS

Selected rating scale factors	*Correlation with total points based on 11 factors*
Plant A:	
Experience (or learning time)	.96
Experience (or learning time) plus hazards	97
Experience (or learning time) plus hazards plus education	.98
Plant B:	
Experience (or learning time)	.93
Experience (or learning time) plus initiative	.95
Experience (or learning time) plus initiative plus responsibility for the safety of others	.96
Plant C:	
Experience (or learning time)	.86
Experience (or learning time) plus hazards	.91
Experience (or learning time) plus hazards plus initiative	.93

In each case the correlations obtained between the most influential three characteristics and the total points determined from all eleven factors were so high that the contribution of the remaining eight factors was negligible.

The facts summarized in Table 14.4 show that *experience* (or learning time) is the most important single factor in determining the total points for the jobs evaluated in all three of the plants. It will also be noted that *hazards* appear among the most important three factors in two of the plants, as does *initiative*. The results further show that three job characteristics give results that, for all practical purposes, are statistically equivalent to those obtained from the use of a more extensive system.

Fig. 14.4 is a graphic representation of the results obtained in Plant A showing the accuracy with which the total points for each job can be predicted from a knowledge of the points assigned on only *three* characteristics—namely, experience, education, and hazards.

The shaded areas in Fig. 14.4 delineate labor grades. From this chart it is apparent that the vast majority of jobs fall in the same labor grade when evaluated on the basis of the stipulated three characteristics as when evaluated on the basis of all eleven factors.

A summary of the effectiveness of the abbreviated three-factor system in Plant A in terms of number of jobs in which the discrepancy in placement between it and the total point rating is one, two, or three labor grades is shown in Table 14.5.

Table 14.5 shows that only seven jobs deviate from their original

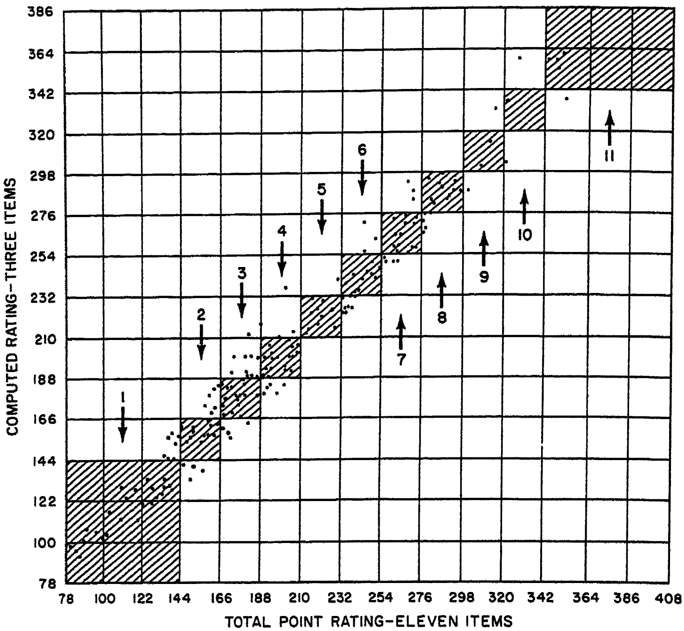


Fig. 14.4. Graph showing ratings computed from three scale items plotted against total point ratings (all eleven items) for 247 jobs in Plant A. The eleven shaded areas define the labor grades designated by the numbered arrows

Table 14.5

DISCREPANCIES BETWEEN TOTAL POINT RATINGS (ELEVEN FACTORS) AND RATINGS COMPUTED FROM THREE FACTORS FOR PLANT A

	Number of jobs by labor grade displacement			
Points of deviation	*Same labor grade*	*Displaced one labor grade*	*Displaced two labor grades*	*All jobs*
0–4	68	9		77
5–9	48	27		75
10–14	20	28		48
15–19	11	21		32
20–24	6	2		8
25–29		4		4
30–34		1	1	2
35–39			1	1
Totals	153	92	2	247

placement by more than 24 points, the approximate range of points covered by one labor grade in this system.

In an investigation of the type summarized above, it is possible that a small number of factors can be extracted from the original set so that for the plant installation under investigation the job evaluations yielded by the complete set of factors will be practically duplicated by the abbreviated system, even though the shortened scale would not work equally well as an abbreviation of the same original system in other plants. Otis and Leukart[19] have therefore formulated a question about the general applicability of a shortened scale. They ask: "How would one go about building an abbreviated scale? There is no way of knowing which one of the shortened scales obtained by Lawshe would best apply to a given plant." Otis and Leukart point out that to decide upon one particular shortened scale that will accomplish the same results as a longer scale in any given plant, the longer scale would first have to be used with all of the jobs and statistical analysis then made to identify the specific factors that will work best as an abbreviated scale for that particular plant installation. They state that the savings in time therefore would not be very great.

To test whether this criticism is valid, Davis and Tiffin[20] selected three factors, each of which appeared in at least two of the three prediction equations developed by Lawshe,[21] and applied this prediction equation to eight entirely different job evaluation installations. Three of the eight installations were NEMA plans, and five were modified point systems. The three factors used were experience, initiative and ingenuity, and hazards. The correlations obtained between total points and points predicted from the three factors are summarized in Table 14.6. For each of the three NEMA installations the three factors used yielded correlations of .94 or higher with the total points from the entire set of factors. For four of the five point systems that were not NEMA plans the corresponding correlations were .91 or above. For only one of eight installations studied, in Company 5, was the correlation too low to justify predicting total points from the abbreviated scale. This one company was a public utility and was using a job evaluation system that included factors quite different from those included in the NEMA plan.

It would thus seem that an abbreviated system does give essentially the same results as a longer system, and that the short system does not need to be developed specially for every plant where it might be applied.

---

19 Otis and Leukart, *op. cit.*, p. 126.

20 M. K. Davis and J Tiffin, "Cross Validation of an Abbreviated Point Job Evaluation System," *Journal of Applied Psychology*, **34** (1950), 225–228.

21 C. H. Lawshe, Jr., "Studies in Job Evaluation: 2. The Adequacy of Abbreviated Point Ratings for Hourly-Paid Jobs in Three Industrial Plants," *Journal of Applied Psychology*, **29** (1945), 177–184.

Salary jobs are ordinarily evaluated by a system that is different from the system used with hourly-rate jobs. The NEMA has a system that includes eleven factors for the evaluation of salary jobs. Lawshe and Maleski[22] have reported a study showing how the evaluations of this 11-factor system can be practically duplicated by an abbreviated system containing only one, two, or three of the original factors. The correlations between results obtained from a reduced number of characteristics and total point values based upon all 11 factors are summarized in Table 14.7.

Table 14.6

CORRELATIONS BETWEEN TOTAL POINTS AND TOTAL POINTS PREDICTED FROM THREE FACTORS IN EIGHT JOB EVALUATION INSTALLATIONS

Company	*No of jobs*	*Correlation*
No. 1 NEMA	341	.96
No. 2 NEMA	126	.95
No. 3 NEMA	605	.94
Non-NEMA Point Systems		
No. 4, 12 factors	61	.98
No. 5, 11 factors	77	.69
No. 6, 10 factors	185	.96
No. 7, 11 factors	273	.96
No. 8, 23 factors	253	.91

In the results for salaried jobs, as in those previously discussed for shop jobs, the experience factor is shown as the most important single characteristic. Also, as in the former instance, ratings based on only three factors are revealed as equivalent to those based on the larger number of factors.

Table 14.7

CORRELATIONS OBTAINED BETWEEN RATINGS ON SELECTED FACTORS AND TOTAL POINT RATINGS FOR APPROXIMATELY 400 SALARY JOBS

Selected scale factors	*Correlations with total points based on 11 factors*
Experience	.93
Experience plus complexity of duties	.96
Experience plus complexity of duties plus character of supervision	.98

[22] Lawshe and Maleski, *op. cit.*

The preceding discussion has been concerned with work on abbreviating the NEMA point systems of job evaluation for shop jobs and salaried jobs. The question may arise whether the same pattern of results would be obtained if another type of point system were similarly analyzed. In answer to this question, Lawshe and Alessi,[23] report a study in which the same type of analysis was made of ratings obtained with a point system differing from the NEMA system in that each job is rated on seven characteristics by means of more and finer categories or degrees for each of the factors, and that the point ratings are translated into "rating factors" by a logarithmic conversion chart for the purpose of assigning monetary equivalents. Table 14.8 summarizes the results obtained from this analysis.

From the correlations summarized in Table 14.8 it is apparent that the stipulated three factors give results practically identical with those obtained from the more elaborate system.

After observing the results of statistical simplification of point systems of job evaluation, the question may be raised as to whether corresponding steps toward simplification of other kinds of job evaluation plans will be equally effective. A similar type of statistical analysis

Table 14.8

CORRELATION COEFFICIENTS BETWEEN RATINGS ON SELECTED SCALE FACTORS AND TOTAL POINT RATINGS OBTAINED WITH A POINT SYSTEM DIFFERING FROM THE NEMA SYSTEM

Selected factors	*Correlation with total points based on 7 factors*
Responsibility	.92
Responsibility plus manual skill	.97
Responsibility plus manual skill plus working conditions	.98

Table 14.9

CORRELATION COEFFICIENTS BETWEEN RATINGS ON SELECTED FACTORS AND TOTAL POINTS OBTAINED WITH THE FACTOR COMPARISON SYSTEM OF JOB EVALUATION

Selected factors	*Correlation with original ratings*
Skill requirements alone	.94
Skill plus working conditions	.97
Skill plus working conditions plus mental requirements	.99

[23] Lawshe and Alessi, *op. cit.*

of the factor comparison system as installed in a paper mill has been reported by Lawshe and Wilson.[24] Here also the results show that a reduced number of job characteristics give results that correlate almost perfectly with the total factor evaluations. These correlations are summarized in Table 14.9.

The fact that fewer than the original number of factors are necessary to give essentially identical results is also shown in a study by Howard and Schutz.[25] These men factor-analyzed a 12-factor salary job evaluation plan and found that a single fundamental factor corresponding to Lawshe's "Skill Demands" accounted for 99 per cent of the variance in total points. Oliver and Winn[26] report similar results. They found that 6 factors of an 18-factor point system plan account for most of the variance of the system. The short 6-factor system classified 62 of 86 salaried jobs in same salary grade as the 18-factor plan and there was a discrepancy of only one salary grade in the classifications of the remaining 24 jobs.

Similar results on this subject have been reported by Myers,[27] who factor-analyzed a point system involving 17 job requirements or characteristics Myers found that 5 fundamental factors accounted for virtually all of the variance of the original 17 characteristics, which strongly suggests that a 5-characteristic system would have achieved the same results as the much longer and more unwieldy plan.

An unpublished study by Scott[28] describes a 6-factor system that has been found to be both reliable and valid. The 6 factors involved are:

1. General Educational Development
2. Specific Job Preparation
3. Physical Demands
4. Working Conditions
5. Job Hazards
6. Supervision

From the results of the several comprehensive studies reported by Lawshe and others summarized above it seems rather clear that job evaluation results statistically equivalent to those derived from more elaborate systems can be obtained from a very much abbreviated plan.

[24] C. H Lawshe, Jr, and R. F. Wilson, "Studies in Job Evaluation. 5. An Analysis of the Factor Comparison System as It Functions in a Paper Mill," *Journal of Applied Psychology,* 30 (1946), 426–434.

[25] A. H. Howard and H G. Schutz, "A Factor Analysis of a Salary Job Evaluation Plan," *Journal of Applied Psychology,* 36 (1952), 243–247.

[26] J. A. Oliver and A. Winn, "An Abbreviated Job Evaluation Plan for Salaried Personnel," *Personnel,* 28 (1951), 225–229.

[27] J. H. Myers, "An Experimental Investigation of 'Point' Job Evaluation Systems," *Journal of Applied Psychology,* 42 (1958), 357–361.

[28] W. E. Scott, Jr., *The Reliability and Validity of a Six-Factor Job Evaluation System,* Ph.D. Thesis, Purdue University, January, 1963.

RELIABILITY OF JOB EVALUATION

Any form of job evaluation is basically an attempt to measure the relative worth of various jobs by a process of individual, pooled, or committee judgments. The ratings obtained (regardless of the system in use) should have a satisfactory reliability, just as the results of any other type of measurement must have a satisfactory reliability, before usable results can be expected. Other things being equal, if two systems have different reliabilities, the one with the higher reliability should always be chosen.

In order to estimate the reliability of the most commonly used point system (the NEMA plan), and also to estimate the reliability of a simplified point system based upon the work summarized in the preceding sections, a series of reliability studies has been conducted and reported by Lawshe and Wilson.[29] In this work, forty job descriptions of commonly known jobs were chosen for evaluation. Twenty men experienced in job evaluation made the ratings. Each job was rated by five men using the NEMA plan, and also by five other raters using a simplified plan that included only four factors. Each rater made his ratings independently. The four factors included in the simplified plan were chosen after factor analyses had been made of several more elaborate systems. These factors were selected to cover as fully as possible the basic characteristics revealed by the several factor analyses.

The results of the reliability investigations of the two systems are summarized in Table 14.10. In this table, the factors used in the NEMA system are listed in column (1), in the order of their obtained reliabilities. The reliabilities, listed in column (2), were obtained by computing the coefficient of correlation between the ratings of each rater and those of every other rater for each factor in the list. The average of the correlations thus obtained was taken as the best estimate of the reliability of a single rater on the factor. In column (3) of Table 14.10 are given the "stepped up" reliabilities, i.e., the values that would be expected if averages of the ratings of 5 raters working independently were used consistently.

The 4 factors included in the simplified system are listed in column (4). Each factor is located to correspond with the approximately equivalent factor of the longer scale.

Columns (5) and (6) give, respectively, the obtained and "stepped up" reliabilities of the factors of the shorter scale. At the bottom of the

[29] C. H. Lawshe, Jr., and R. F. Wilson, "Studies in Job Evaluation: 6. The Reliability of Two Point Rating Systems," *Journal of Applied Psychology*, 31 (1947), 355–365.

Table 14.10

RELIABALITY DATA ON THE NEMA JOB EVALUATION SYSTEM AND A SIMPLIFIED POINT SYSTEM INVOLVING FOUR FACTORS

NEMA system			*Simplified system*		
(1) *Factor*	(2) r	(3) r_s	(4) *Factor*	(5) r	(6) r_s
Experience	.82	.96	Learning period	.86	.97
Initiative and ingenuity	.78	.95			
Education	.77	.94	General schooling	.79	.95
Responsibility for safety of others	.54	.85			
Working conditions	.54	.85	Working conditions	.61	.89
Responsibility for work of others	.51	.84			
Physical demand	.47	.82			
Responsibility for equipment or processes	.41	.78			
Responsibility for materials	.40	.77			
Mental or visual demand	.37	.75			
Unavoidable hazards	.34	.72	Job hazards	.51	.84
Total Points	.77	.94	Total Points	.89	.98

table are given the obtained and "stepped up" reliabilities of the total scales.

It will be noted in Table 14.10 that the factors of the shorter scale are all higher in obtained reliability than the corresponding factors of the NEMA scale, and this is true whether one considers the obtained reliabilities or the "stepped up" reliabilities. It will be further noted that the reliability of the total shorter scale is higher than the corresponding reliability of the longer scale.

The greater reliability of the shorter scale may be due in part to the use of more carefully worded descriptions of the degrees of the various characteristics, rather than to the mere fact that a shortened scale is being used. Familiarity with the scale and the experience with its use will not explain the results, because the majority of raters using the NEMA scale had had practical experience with it, whereas *none* of the raters had ever seen the shortened scale before the experiment.

Table 14.10 also shows that the obtained reliability of the total NEMA point ratings is lower than that of several of the individual factors in this system. In the case of the simplified system, the total points resulted in a higher reliability than that obtained for any specific factor.

Further data on the reliability of a shortened system have been reported by Scott.[30] Using the 6-factor system that he developed, 15 jobs were independently evaluated from the corresponding 15 job descriptions

[30] Scott, *The Reliability and Validity of a Six-Factor Job Evaluation System,* Ph D. Thesis, Purdue University, January, 1963.

by several committees, each committee consisting of five men. Some of the committees consisted of men who had had experience with other job evaluation plans and some were made up of men with no previous experience except for a two-day instructional session on the nature and purpose of job evaluation. The reliability of the job evaluations obtained are summarized in Table 14.11. These reliability coefficients were obtained from average correlations between results turned in by the committees with similar backgrounds. The results summarized in Table 14.11 show that committees with only minimal instruction make evaluations that are just as reliable as do committees that have had rather extensive experience with job evaluation. The table further shows that the results obtained from committees of both types are very reliable, i.e., the jobs are evaluated at very similar levels by committees of both types. The correlation between evaluations made by an "expert" committee and those made by an inexperienced committee using the six-factor system was found to be .96. Further results showed that different committees (experienced or inexperienced) did not differ significantly in rating level. Total points assigned to the jobs by one committee were not significantly higher or lower than those assigned by several other committees rating the same jobs from the same job descriptions.

Table 14.11

THE RELIABILITY OF A 6-FACTOR JOB EVALUATION SYSTEM AS DETERMINED FROM TWO KINDS OF COMMITTEES USING THE PLAN

	Reliability	
(1)	(2) *Experienced committees*	(3) *New committees*
General Educational Development	.91	.95
Specific Job Preparation	.96	.96
Physical Demands	.80	.97
Working Conditions	.79	.92
Job Hazards	.81	.94
Supervision	.94	.99
Total Points	.96	.97

The two investigations summarized above (Lawshe and Wilson[31] and Scott[32]) each made use of 5-man committees. The reliability of the job evaluations obtained clearly shows that highly reliable results (.94 or higher) can be obtained with a committee of that size. On the other hand, some results published by the United States Air Force, in which 50 Air Force Specialities (jobs) were evaluated by student officers attending the

[31] Lawshe & Wilson, *op. cit.*
[32] Scott, *op. cit.*

Command and Staff School,[33] indicate that 20 raters were needed to achieve a reliability of .93 or above. Whatever may account for this difference, the results obtained by Lawshe and Wilson and by Scott show that one does not need 20 civilian raters to achieve reliable results in evaluating factory jobs with a system designed for use in such organizations.

The study by Chesler,[34] referred to previously, also deals with reliability of job evaluation. In this study seven raters from seven different companies each evaluated the same 35 standard jobs from job descriptions. A standard job evaluation manual with 12 factors was used by all raters. The evaluations made by each rater were correlated with those from every other rater. The correlation coefficients ranged from .93 to .99, with an average of .97. These results indicate that different raters, using the same manual and evaluating the same jobs, agree very closely with each other in the job evaluations obtained. Chesler's findings also indicate that: (1) The ratings from a shortened system correlate very highly with the ratings of longer systems, (2) The findings are essentially the same for shop jobs and salaried jobs, (3) The effect of shortening the system is the same for point systems and for the factor comparison system, and (4) Ratings obtained from any standardized point system have a satisfactory degree of reliability.

In the light of these findings—that job evaluation systems consisting of only a few factors give essentially the same results as systems involving many more factors, and that a shortened system is at least as reliable as a longer system—it would seem advisable in the future to make greater use of systems considerably shorter than many that are in current use.

SUMMARY

Perhaps the primary implication of this discussion of job evaluation is in terms of emphasizing the utility to management of a systematic, quantitative analysis of information relating to job evaluation systems that are in actual use, or that are being considered for adoption. Such an analysis can provide management with information that can be useful in making more adequate decisions regarding job evaluation programs.

The measurement, quantification, and analysis of data are, of course, central to the principal point of view that the writers have attempted to emphasize throughout this text. In the field of personnel management,

[33] R. E. Christal, J. M. Madden, and F. D. Harding, "Reliability of Job Evaluation Ratings as a Function of Number of Raters and Length of Job Descriptions," WADD-TN-60-257, Personnel Laboratory, Wright Air Development Division, Air Research and Development Command, U. S. Air Force, Lackland Air Force Base, Texas.

[34] Chesler, *op. cit.*

as well as in other areas, there are no substitutes for solid, hard facts when making decisions. The systematic application of some of the methods and procedures discussed in this book offers considerable promise of providing pertinent facts about some of the human problems in business and industry that can be used in making decisions directed toward their solution.

IV

Work Situation

The last section dealt with the organizational and social context of human work. This section is concerned more with the nature of the physical environment, the physical equipment that is used, the nature of the job activities, and related aspects of the work situation. Here, again, these are attributes of the work situation which have some effect upon the adequacy with which human work is carried out and upon human welfare. Appropriate attention to these considerations can contribute to the more effective utilization of human effort. Particular mention might be made of the interest in recent years in what is referred to as human factors engineering. This area is concerned

with the design of equipment and physical facilities with particular attention to considerations of human abilities and limitations. In a world of increasing technological developments, this is an area of increasing importance.

15

Working Conditions

As indicated earlier, human work behavior can be influenced by situational variables, including those to which we commonly refer as *working conditions.* There are really three groups of working conditions: those of a physical nature (illumination, noise, atmospheric conditions, etc.), those relating to time (hours of work, rest pauses, etc.), and those relating to the social situation within which the individual works. Since the social aspects of the work situation were discussed earlier, we shall be concerned here with the physical aspects of working conditions and time. In any endeavor in which the effects of working conditions are to be investigated, it is, of course, necessary to utilize relevant criteria as the basis of comparison or measurement.

CRITERIA IN EVALUATING WORKING CONDITIONS

In Chapter 2 it was suggested that criteria that are relevant to industrial psychology can generally be grouped into the following classes: performance, subjective, physiological, and (in some cases) accidents. Since accidents will be discussed in Chapter 18, this criterion will not be

discussed here. All of these can be useful in comparing the effects of working conditions upon people.

Performance Criteria A fairly common type of criterion used in evaluating working conditions is some measure of performance. While work output is perhaps the most common performance measure, other performance criteria sometimes are appropriate, especially those relating to basic human processes. Among such criteria are visual performance, motor performance, and mental performance. Where the effects of duration of work periods are being studied, performance decrement may be used as a criterion.

Physiological Criteria Human work is accomplished by certain physiological processes. As a person performs work, especially physical work, there typically are various changes in his physiological condition. If the work is severe enough or long enough, an individual's physical ability to perform the task deteriorates.

The fact of reduction in physical ability to perform a task may be readily demonstrated by simply physiological experiments in which a muscle or muscle group is caused to undergo a simple rhythmic contraction. Thus, if one squeezes a coiled spring at intervals of two seconds, he will find that the squeezes become less and less powerful until finally, if the task is maintained for sufficient time, only very small changes, or no changes at all, in the spring can be noticed. The energy used in the execution of a task of this sort, or in any muscular task, comes from potential energy that is stored in chemical form in the muscles. As this energy is expended, the muscles become less and less able to perform their task. This reduction in potential energy available in the muscles may be thought of as *physiological fatigue*. The phenomena of fatigue, from this point of view, probably depend upon two factors, namely, (1) the consumption of the contractile material or of the substances available for the supply of potential energy to this material, and (2) the accumulation of waste products of contraction. These waste products may be thought of as the chemical result of the muscular activity. Among them lactic acid is of particular importance, because it is known that fatigue may be artificially induced in a muscle by feeding the muscle with a dilute solution of lactic acid.

Various physiological measures are used as indexes of physiological change in the organism. These include heart rate, blood pressure, oxygen consumption, breathing rate, blood composition, and electrical resistance in the skin. A measure that has been studied rather extensively by Brouha[1] is the heart rate recovery curve. This is obtained by counting the pulse rate at one-minute intervals during the first three minutes of

---

[1] L. Brouha, "Fatigue—Measuring and Reducing It," *Advanced Management*, 19 (January, 1954), 9–19.

the recovery period after the termination of work and while the subject is sitting quietly. The curve based on such readings indicates the actual value of the pulse and the rate of recovery toward resting level. Fig. 15.1

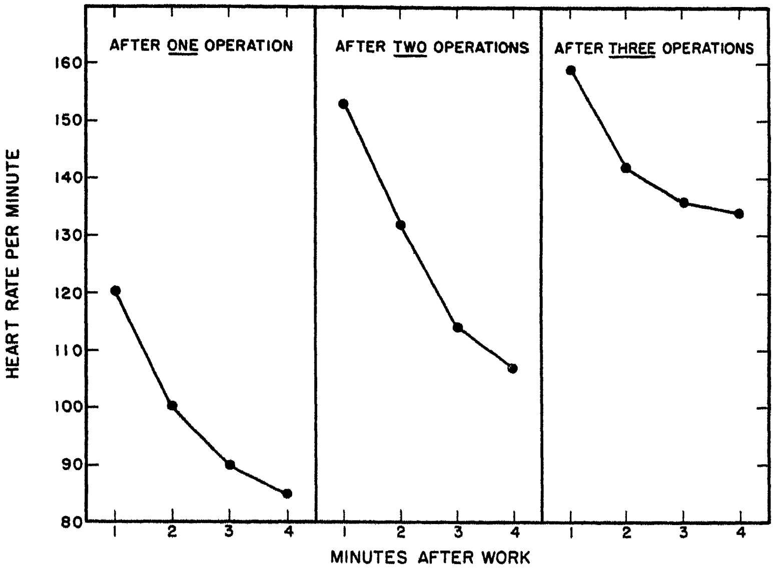


Fig. 15.1. Change in heart rate recovery curves after one, two, and three repetitions of a strenuous physical operation. (From Brouha, *op. cit.*)

shows the heart rate recovery curves resulting from one, two, and three repetitions of a strenuous physical operation. The differences in the levels of these curves are indicative of differences in the physiological stress of performing the operation once, twice, and three times.

Physiological changes are greatest, of course, with predominantly physical, as opposed to predominantly mental, work activities. While there are physiological changes that occur during mental activities, these changes are quite small, however, and to date have been of limited practical value as criteria of the effects of mental work.

Psychological Criteria It is a matter of common experience that work usually may be accompanied by some subjective (psychological) reactions to it. There are various dimensions of such reactions. Among the most common subjective reactions are those of *boredom* and *subjective fatigue.* What we commonly call boredom usually is associated with work that is intrinsically uninteresting to the individual performing it. Most frequently, boredom is associated with jobs that are repetitive or relatively simple. It should be pointed out, however, that it is the individual's

reaction to the job that would cause him to describe a job as "boring." No job is, or can be, intrinsically boring, since boredom is an attitudinal reaction of the individual. What we will call *subjective fatigue* (also called *psychological fatigue*) refers to the sensations of boredom, of weariness, of being "beat"—feelings that we all experience at one time or another. Still another possible type of psychological reaction to work situations is that of psychological stress. Stress is most often viewed as the consequence of competition between two incompatible tendencies, one of which typically has a positive valence and the other a negative valence.[2] Thus, situations in which frustrations, conflict, fear, or anxiety can arise might give rise to psychological stress. It has been emphasized by Vogel, Baker, and Lazarus,[3] however, that stress can occur only under conditions of motivation, when there is some threat to the attainment of some goal.

Generally speaking, if we wish to ascertain for an individual his level of boredom, subjective fatigue, or psychological stress, we have to ask him. (It should be noted, however, that physiological concomitants of some psychological states can be detected, especially when the individual is emotionally aroused.) The measurement of the subjective reactions of people raises some interesting problems in semantics, especially in the meaning of words. Some words have different meanings for different people. Therefore, in expressing differing degrees of boredom or subjective fatigue, one has to use words that have similar meanings to the person expressing his feelings and the person to whom he is expressing them.

In this connection, McNelly[4] used the Thurstone scaling method (described previously in Chapter 11) in the scaling of words relating to subjective feelings of fatigue. A scale was constructed consisting of nine words which had been judged rather reliably in terms of subjective fatigue, and which represented varying "degrees." Actually, two forms were developed. One of the forms is given in Table 15.1. This shows each of the nine words or phrases, its scale value (actually the mean of the categories in which the several judges classified it), and its standard deviation (which shows the variability of those judgments). The results of an experimental laboratory study with this scale suggest that it may have some potential utility as a criterion of the feelings of fatigue that people experience.

---

[2] W. D. Chiles, *Psychological Stress As a Theoretical Concept,* USAF Wright Air Development Center, Technical Report 57-457, July, 1957.

[3] W. Vogel, R. W. Baker, and R. S. Lazarus, "The Role of Motivation in Psychological Stress," *Journal of Abnormal and Social Psychology,* 56 (1958), 105–116.

[4] G. W. McNelly, *The Development and Laboratory-Validation of a Subjective Fatigue Scale,* Ph.D. Thesis, Purdue University, August, 1954.

Table 15.1

WORDS AND PHRASES IN A SCALE OF "SUBJECTIVE FATIGUE" SHOWING MEAN SCALE VALUES AND STANDARD DEVIATIONS OF JUDGMENTS

Word or phrase	*Mean scale value*	*Standard deviation*	*Rounded scale value*
About to fall over	1.1	.30	1
Fagged	2.2	.36	2
Let down	3.2	.51	3
A little tired	3.8	.40	4
Average	5.0	.00	5
Fairly well	5.9	.57	6
In gear	7.0	.63	7
Very good	7.9	.54	8
Terrific	8.9	.30	9

Source: *McNelly*, op. cit.

Discussion of Criteria Since performance, physiological, and psychological criteria might be used in investigations of working conditions (hours of work, temperature, noise, etc.), one might wonder about the relationships among these various criteria. In general, these criteria are not highly correlated. There are, however, certain types of situations in which there are significant relationships between and among such criteria. For example, in the case of heavy physical work continued over a period of time, it would be expected that physiological changes would take place, and that performance also would deteriorate; subjective fatigue might also develop. But it is possible for performance to deteriorate without physiological deterioration, or for work performance to remain at a fairly high level (say, due to strong motivation) in the presence of subjective fatigue.

In general, then, one criterion (such as performance) cannot be used as an index of another (such as a physiological measure). This lack of general relationship then suggests that in individual investigations of working conditions or of the effects of continued work, that criterion be used which is most relevant to the nature and purpose of the investigation. For example, in studying the effects of temperature on people, one might appropriately use some physiological measure. In studying the effects of rest pauses, however, one would more likely use an index of performance. But while one particular criterion might be most relevant in a given situation, it is frequently possible to use two or more. For example, while in industrial organizations performance criteria frequently are of particular interest, industry should not be unmindful of

the physiological and psychological effects of work and working conditions.

ILLUMINATION

The ability to make visual discriminations is the consequence of three general classes of variables: individual differences, the nature of the visual task, and illumination. Individual differences in visual skills were discussed in Chapter 6, and so will not be dealt with here.

The Nature of the Visual Task The "task" variables that affect the ability to make visual discriminations include: the brightness contrast of the "details" to be discriminated against their background; the size of the details to be discriminated; and the time available for seeing. It is, of course, sometimes possible to modify visual tasks, especially the visual objects to be viewed, in order to make them more readily discriminable. This is possible, for example, in the design of visual displays, the size and design of labels and printed material, etc. Some aspects of this will be discussed in Chapter 16.

A couple of sets of data relating to task variables will be given to illustrate the effect of such variables on visual performance. One example, from a much broader study by Blackwell,[5] deals with the relationship between brightness contrast and visual performance. While the details of the study will not be discussed, the results are shown in general terms in Fig. 15.2. An example of low brightness contrast would be gray printing on a slightly lighter gray paper. An example of high contrast would be black printing on very white paper. Another example, also from Blackwell,[6] shows the interaction between viewing time and brightness contrast, as related to illumination level. The data, shown in Fig. 15.3, can be interpreted as follows: For any given time curve (such as one second viewing time), the various combinations of illumination level and brightness contrast produce the *same level* of visual discriminability. (The level of discriminability depicted in this figure was 50 per cent detection of the particular visual stimuli that were used.) For any given level of illumination, it can be seen that if viewing time is *decreased,* the brightness contrast must be *increased* in order to maintain the same level of visual discriminability.

Level of Illumination It is frequently the case, of course, that the nature of the visual task (brightness contrast, size of detail, etc.) cannot be modified; it is predetermined by the nature of the work situation. This is the case, for example, of visual inspection of products, the observation of materials in process, or the observation of objects in one's

[5] H. R. Blackwell, "Development and Use of a Quantitative Method for Specification of Interior Illumination Levels," *Illuminating Engineering,* 54 (No. 6, June, 1959), 317–353.

[6] *Ibid.*

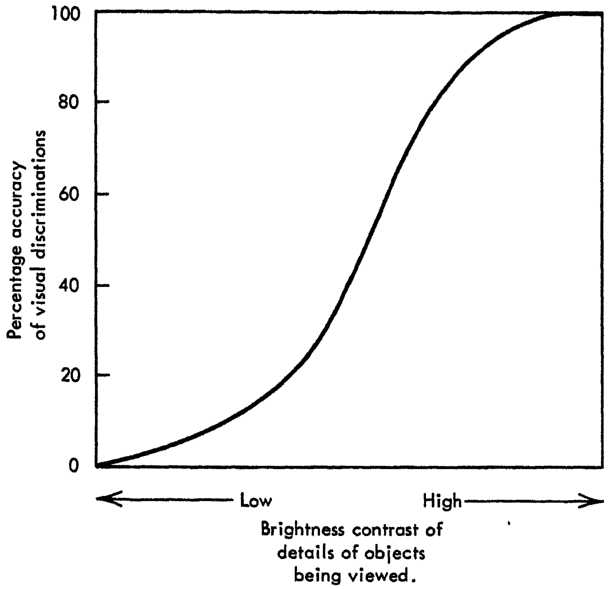


Fig. 15.2. Generalized relationship between brightness contrast of visual detail and accuracy of visual discriminations. (Adapted from Blackwell, *op. cit.*, Fig. 4.)

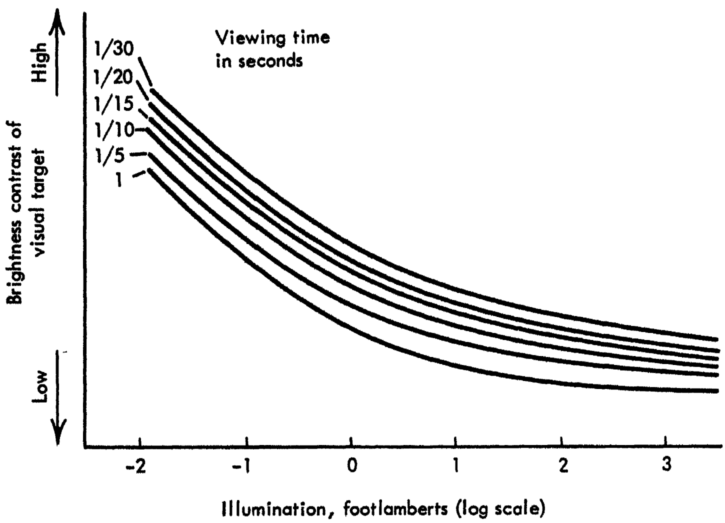


Fig. 15.3. Illumination and brightness contrast relationships for each of six viewing times. The curve for any viewing time represents the combinations of illumination and brightness that are required for equal visual discriminability. (Adapted from Blackwell, *op. cit.*, Fig. 33.)

environment. Where this is the case, and especially where the details are difficult to detect, the illumination available is especially critical. And even where the task variables can be controlled, the illumination should be set at a level that insures reasonably satisfactory visual performance

The effects of varying levels of illumination on performance have been demonstrated in various work-situation and laboratory studies. A generalized pattern of relationship is shown in Fig. 15.4. The specific pat-

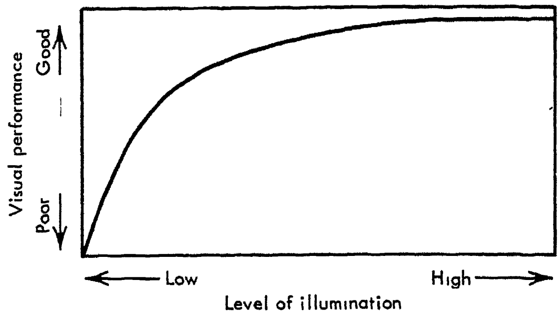


Fig. 15.4. Generalized relationship between level of illumination and visual performance. As illumination increases from a very low level, performance increases, first sharply, and then more slowly, ultimately it tends to level off.

tern would vary with the task and the criterion used, but in the case of most visual tasks, starting from a low level of illumination, visual performance first rises rather sharply, then increases more gradually, and then tends to level off.

Setting Illumination Standards Over the years there have been differences of opinion regarding the standards by which illumination levels for various work tasks should be evaluated. In recent years, however, the extensive research program by Blackwell[7] is coming to be accepted by many (including the Illuminating Engineering Society) as providing the base for setting such illumination levels. The details of his research methods, and the way in which his research results are converted into recommended illumination levels, will not be given here. The bases for the recommendations are discussed by Crouch,[8] and the recommendations as such are reported by a committee of the Illuminating Engineering Society.[9] For purposes of providing a base for recommendations for specific job activities, five categories of "seeing tasks" were established, each of

[7] Blackwell, *op cit.*

[8] C. L. Crouch, "New Method of Determining Illumination Required for Tasks," *Illuminating Engineering*, 53 (No. 8, August, 1958), 416–422.

[9] "Recommendations for Quality and Quantity of Illumination," *Illuminating Engineering*, 53 (No. 8, August, 1958), 422–432.

these representing tasks which, in general, have brightness contrast values that require relatively similar levels of brightness (reflected illumination). These categories are given below:

Category of seeing task	*Require brightness in footlamberts*
Most difficult	420 up
Very difficult	120–420
Difficult	42–120
Ordinary	18–42
Easy	Below 18

A footlambert can be thought of as the brightness of *reflected* light that is equivalent to a perfectly diffusing, perfectly white surface which is illuminated by one footcandle of light. The number of footcandles (fc) that would be required to produce any specified level of footlamberts would depend upon the reflectance of the surface. Taking, as an example, a "difficult" task that would require, say, 100 footlamberts, if the surface reflected 80 per cent of the light, the footcandles required would be $100 \div .80$, or 125 fc. If the surface reflected only 25 per cent, the footcandles required would be $100 \div .25$ or 400 fc. Thus, for any given level of task difficulty, the actual footcandle requirements will depend very largely on the reflectance of the surfaces of visual attention. This argues for the use of fairly light work surfaces such as desks, work tables, etc. On the basis of the procedures developed by Blackwell, illumination levels have been recommended for various types of tasks.[10] A few examples of these recommendations are given below:

Task	*Recommended illumination, footcandles*
Very difficult inspection	500
Proofreading	150
Regular office work	100
Wrapping and labeling	50
Dishwashing	30
Loading (materials handling)	20
Hotel lobby	10

Discussion It should be added that there are other important aspects of illumination besides the level. One of these is the location of luminaries; these should be installed in such positions as to minimize glare. Further, in addition to providing the appropriate level of illumination at the work area, there should be a reasonable amount of general illumination surrounding the work area. And in certain special situations (as in some inspection operations) the spectral characteristics of the illumination sources need to be given particular attention.

[10] *Ibid.*

ATMOSPHERIC CONDITIONS

Different kinds of variables can be thought of as being aspects of our atmosphere. Besides temperature and humidity there are also the factors of air flow, barometric pressure, composition of the atmosphere, and sometimes toxic conditions. There is also the factor of temperature of objects in the environment; this is not strictly an "atmospheric" condition but it certainly relates to this subject. Our discussion will deal largely with the more common aspects of the atmosphere, especially temperature and humidity.

The Heat-Exchange Processes The metabolic processes of the body result in the generation of heat, some of which the body normally has to dispose of. In disposing of such heat, the body is continually attempting to maintain a thermal equilibrium with its environment. The heat-exchange process is the method by which the body attempts to achieve this balance.

There are four ways in which this heat exchange takes place. Convection is the transmission of heat by a fluid that occurs when there is a temperature difference between the object and the gas. In the case of people, the body typically transmits heat to the air (which is technically a fluid), although when the air temperature gets above body temperature the transmission is reversed. *Evaporation* is another method of heat exchange; this consists primarily of evaporation of perspiration, and to some degree of vapor that is exhaled from the lungs in breathing. *Radiation* is the process of transmission of thermal energy between objects such as between the sun and the earth. Such transmission occurs either with or without an atmosphere if there is a difference between the temperatures of two objects, the warmer object losing heat to the cooler. Usually people transmit heat to *other* objects by this method, but occasionally a person may be in a situation where the objects transmit heat to him, such as from boilers or heated metal. *Conduction* is the transmission of heat by direct contact, such as with chairs or the floor. Our clothing usually so insulates us that this is a very unimportant method of heat exchange.

Conditions That Affect Heat Exchange The amount of heat that is dissipated from the body by the different methods varies with the atmospheric conditions. With high humidity, for example, less heat can be lost by evaporation than with low humidity. Some indication of the relative importance of different methods of heat loss, in relation to atmospheric conditions, is reflected in examples given by Winslow and Herrington.[11] These are shown in graphic form in Fig. 15.5. For each

[11] C. E. A. Winslow and L. P. Herrington, *Temperature and Human Life* (Princeton, N. J.: Princeton University Press, 1949).

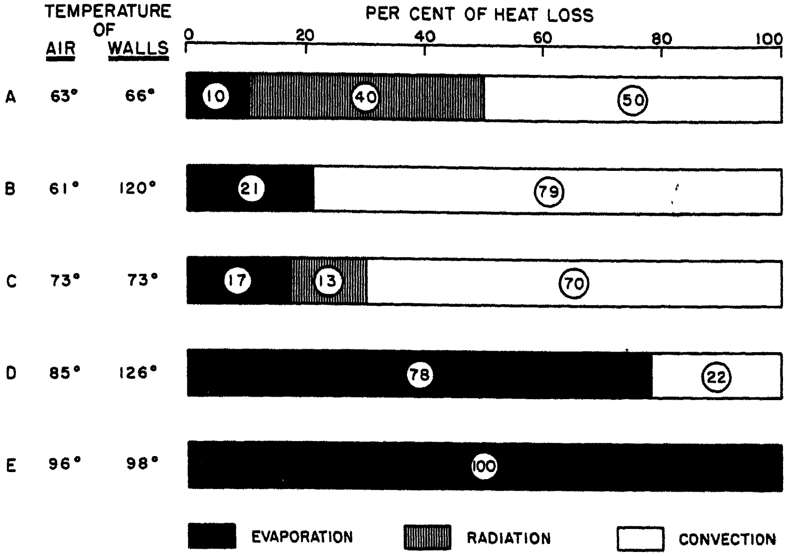


Fig. 15.5. Per cent of heat loss to environment by different methods under different combinations of air and wall temperature. (Adapted from Winslow and Herrington, *op. cit.*, p. 55.)

of five combinations of *air temperature* and *wall temperature* there is shown the *per cent* of the total heat loss that occurs by each of three methods: evaporation, radiation, and convection.

With high wall temperatures (such as B, D, and E, with wall temperatures of 120, 126, and 98) the loss of heat by radiation is not possible; the body would be *gaining* heat by this process. With air temperatures rather low (as in A, B, and C) the important role of convection is seen in the figure. With high air temperatures (D and E) the heat-loss burden is thrown almost exclusively on the evaporation process.

Fig. 15.5 does not show the effects of humidity, since humidity in the study was moderate. If one includes a high humidity level with high temperature, evaporation also is limited, so the body cannot lose the amount of heat it should lose. The result is extreme discomfort, and reduction in the physical ability to perform tasks.

Fig. 15.6 shows the approximate upper limits of tolerance for heat loss by evaporation for three conditions: resting with air movement, resting without air movement, and working. Combinations of temperature and humidity to the right of a line would preclude evaporative heat loss.

Effects of Atmospheric Conditions on Performance While there are now relatively fewer jobs that are performed under adverse atmospheric

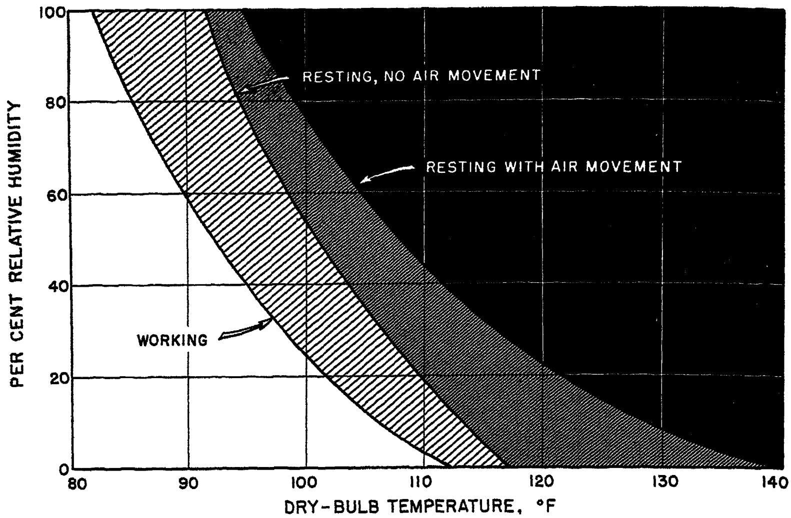


Fig. 15.6. Approximate upper limits for heat loss by evaporation Each line shows the limit, in terms of temperature and relative humidity, that typically prevents evaporative heat loss for the three conditions. Data are for resting or working nude subjects, for clothed subjects, tolerance limits would be noticeably lower. (Adapted from C. E. A. Winslow *et al*, "Physiological Influence of Atmospheric Humidity," Second Report of the ASHVE Technical Advisory Committee on Physiological Reactions, Transactions of the ASHVE, 1942, 48, 317–326.)

conditions than in the past, there are still many jobs in which the workers are exposed to atmospheric extremes of heat or cold.

Hot conditions. The level of heat that can be tolerated without adverse performance effects is, in part, the consequence of the type of work. In the case of heavy physical work, temperatures generally should be somewhat lower than in the case of light work or sedentary work, because of the increased metabolic rate brought about by heavy work.

Some evidence of the effects of high temperatures upon physical work comes from a study by Mackworth [12] in which a heavy pursuitmeter task was carried out by subjects under conditions of various combinations of effective temperatures. (Effective temperature takes into account both temperature and humidity; the numerical value is that of the temperature of still air in combination with 100 per cent relative humidity, which would induce an identical temperature.) The task involved considerable

[12] N. H. Mackworth, *Researches on the Measurement of Human Performance*, Medical Research Council, Special Report Series 268, H. M. Stationery Office, London, 1950.

physical effort in operating a weighted lever in a pursuit task. The results are shown in line *a* of Fig. 15.7. The increasing error rate with increasing temperatures is quite clear. (Later mention will be made of line *b*.)

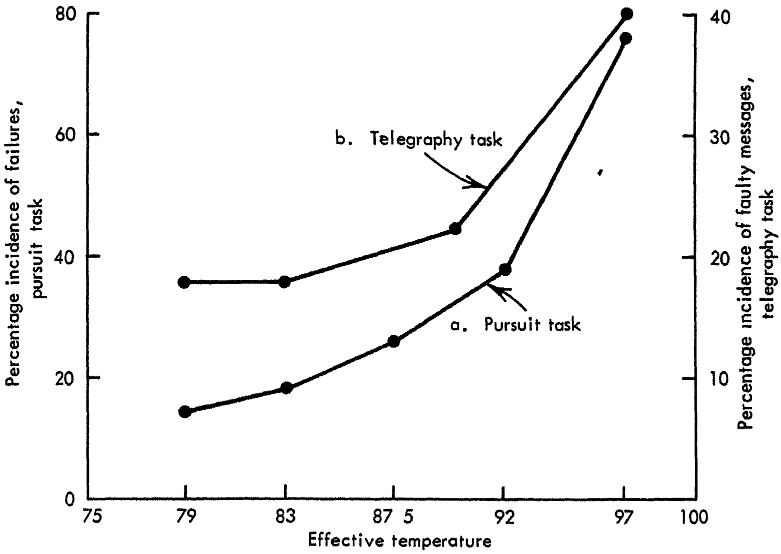


Fig. 15.7. Relationship between effective temperature and performance on a heavy pursuitmeter task and on a telegraphy task. (Adapted from Mackworth, *op. cit.*, Figures 51 and 52, permission H M. Stationery Office, London.)

With respect to the effects of high temperatures on mental activities, there are some conflicting research findings. Chiles,[13] for example, reports no appreciable decrement in performance on a fairly complex mental task over a range of effective temperatures from 76 to 91 degrees. On the other hand, Pepler [14] reports deterioration in a mental task for 5° increments between 76° and 91° effective temperature. Chiles, however, questions his findings on statistical grounds. Mackworth [15] also reports noticeable performance decrement within this general range. In particular, he reports increases in errors in both a wireless telegraphy test and a coding task. The results of the telegraphy test are shown in line *b* of Fig. 15.7.

There is some ambiguity with respect to the upper limits of tem-

[13] W. D. Chiles, *Effects of High Temperatures on Performance of a Complex Mental Task*, USAF, WADC Technical Report 58-323, July, 1958.

[14] R. D. Pepler, *A Complex Mental Task with Varying Speed Stress*, Applied Psychology Research Unit, Cambridge, England, Report No. A.P.U. 196/53, 1953.

[15] Mackworth, *op. cit.*

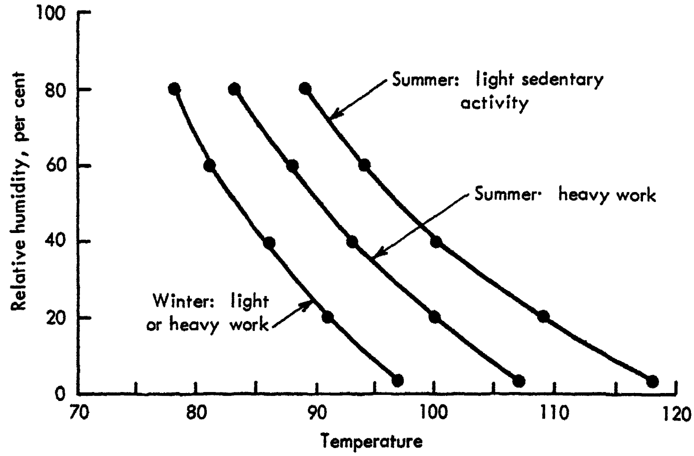


Fig. 15.8. Temperature and humidity combinations that can be tolerated daily for various types of work by healthy, acclimatized men wearing warm weather clothing. These data are for air movement of 15–25 ft/min. (Adapted from report of Committee on Industrial Ventilation, "Thermal Standards in Industry," *Year Book of the American Journal of Public Health* (May, 1950), p. 133.

peratures that can be tolerated for various types of work. Some indication of such limits, however, comes from a report of a Committee on Industrial Ventilation,[16] and are given in part in Fig. 15.8. These data are for healthy, acclimatized men. For persons who are not acclimatized, or who are not in reasonably good physical condition, these limits should be lowered. The data given in Fig. 15.8 are, of course, upper limits. Optimum ranges are considerably lower. In this connection, Baetjer,[17] in summarizing various studies, gives the following estimates of desirable conditions for the types of activities in question:

Activity	*Desirable dry-bulb temperature, F°*
Light sedentary work: winter	68–73
Light sedentary work: summer	75–80
Moderate hard work	65
Strenuous work	60

In these ranges of temperatures, relative humidity is not a particularly important variable.

The four atmospheric variables that are particularly pertinent to

[16] Committee on Industrial Ventilation, *op. cit.*

[17] Anna M. Baetjer, "Light, Temperature, Humidity," *Industrial Medicine*, 13 (1944), 111–112.

comfort and performance in everyday living are temperature, humidity, radiation (the temperature of surrounding objects such as walls and ceilings), and air flow. The first three have been discussed above in one context or another. Air flow is most important in warm or moderately hot environments, since it exposes the body to changing air, thereby speeding up convection and evaporation heat losses. At temperatures up toward 100°, and above, however, its effect is markedly reduced, and, in fact, in combination with high humidities (which limit evaporation) air flow may even reduce heat loss. In warm environments below such extremes, the effect of air flow is cooling, as we all know from sitting in front of an electric fan on a warm day.

Cold. Exposure to cold generally is accompanied by a number of physiological changes, including the vasoconstriction of the peripheral blood vessels, thus reducing the flow of blood to the surface of the skin, resulting in reduced skin temperature. This response has the effect of minimizing heat loss. In studies of the effects of cold on psychomotor tasks, it has been found that task performance is directly related to this reduction in hand skin temperature. In particular, Clark[18] found that the critical skin temperature is in the 55° to 60° range. Manual performance is not affected by skin temperatures above 60°, but decrements may be expected with skin temperatures below 55°.

While there have been very few studies of the effects of cold on higher mental processes, available evidence suggests that such processes are not markedly affected by cold.

Comfort Zones While there are, of course, individual differences in relative comfort under different temperatures, the maximum percentage of individuals feel comfortable in the summer at about 71° effective temperature, and in the winter at about 68°.[19] Variations of a degree or two either way from these would still result in fairly high percentages of people feeling reasonably comfortable.

A great deal has been written concerning the various effects of noise in industry, especially on people's work performance and hearing. Before considering these effects, however, we should mention briefly the measurement of sound.

There are two primary characteristics of sound, namely, *frequency* and *intensity*. The frequency of sound is based on the number of vibra-

[18] R. E. Clark, *The Limiting Hand Skin Temperature for Unaffected Manual Performance in the Cold,* Quartermaster Research and Engineering Command, Natick, Mass., Technical Report EP-147, February, 1961.

[19] American Society of Heating, Refrigerating and Air-conditioning Engineers (ASHRAE), *Guide and Data Book,* 1963, Chap. 8.

tions per second of the noise source. Middle C on the piano, for example, has a frequency of 256 cycles per second (cps); each octave has twice the frequency of the octave below it. Most sounds in our environment, however, do not consist of pure tones, but rather consist of combinations of frequencies. Noise has sometimes been characterized as "unwanted" sound.

Intensity is measured by the decibel (db). Fig. 15.9 illustrates the intensities, in decibels, of various noise sources and environments.

Effects of Noise on Performance The evidence from noise studies indicates that noise does not *generally* cause deterioration in work performance. In fact, Broadbent [20] states that a review of available reports shows no experiments in which there have been statistically significant effects of noise at noise levels less than 90 decibels, although there have been some studies of noise levels at about 90 decibels that have shown significant decrements. (It should be noted, however, that there is no proof that lower levels will *not* impair performance.)

Broadbent [21] points out further that there is no clear cut and obvious distinction between tasks on which performance *is* affected and those on which performance is *not* affected. There are some inklings, however, that suggest that simple paper-and-pencil tasks and sensorimotor tasks seem relatively insensitive to the effects of noise, and on the other hand that monitoring types of tasks may be adversely affected. It also seems that performance on complex mental tasks is susceptible to decrement in the presence of high noise levels.

The general lack of demonstrated adverse effects of noise in the case of laboratory studies may be attributable in part to strong motivation on the part of laboratory subjects. In the case of work situation studies, adaptation to noise probably plays a part, plus the basic problem of studying the effects of a single variable (such as noise) in a dynamic work situation in which other influences may also affect performance.

There is, however, some suspicion that, in the case of certain kinds of tasks, continuous work over a period of hours under high noise levels may take its toll in terms of work performance. In particular, this seems to be the case with paced (as opposed to unpaced) tasks that require extreme speed in responding to relevant stimuli (as in machine operation, inspection of items on a conveyor belt, etc.). Some evidence of this comes from a study by Broussard *et al.*[22] in which choice reaction time was measured under conditions of 90-decibel noise and quiet (45 decibels)

[20] D. S. Broadbent, "Effects of Noise on Behavior," Chapter 10 in C. M. Harris (Ed.), *Handbook of Noise Control* (New York: McGraw-Hill Book Company, 1957).

[21] *Ibid.*

[22] I. G. Broussard *et al.*, *The Influence of Noise on the Visual Contrast Threshold*, Army Medical Research Laboratory Report 101, November 6, 1953.

Noise Source	Decibels	Environment
Hydraulic press (3')		
Large pneumatic riveter (4')	130	
		Boiler shop (maximum level)
Pneumatic chipper (5')		
	120	
Multiple sand-blast unit (4')		
Trumpet auto horn (3')		Jet engine test control room
Automatic punch press (3')		
	110	
Cut-off saw (2')		Woodworking shop
		Inside DC-6 airliner
Annealing furnace (4')		Weaving room
	100	
Automatic lathe (3')		
		Can manufacturing plant
Heavy trucks (20')		Inside Chicago subway car
		Inside motor bus
	90	
10-hp outboard (50')		Inside sedan in city traffic
Small trucks accelerating (30')		
	80	
Light trucks in city (20')		Office with tabulating machines
Autos (20')		Heavy traffic (25' to 50')
	70	
		Average traffic (100')
Conversational speech (3')		Accounting office
	60	
	50	Private business office
		Light traffic (100')
		Average residence
	40	
	30	Broadcasting studio (speech)
		Broadcasting studio (music)
	20	Studio for sound pictures
	10	
Threshold of hearing – young men 1000 to 4000 cps	0	

Fig. 15.9. Typical over-all sound levels for various noise sources and environments; distances in feet are given where appropriate. (Adapted from A. P. G. Peterson and E. E. Gross, Jr., *Handbook of Noise Measurement* (5th ed.), General Radio Company, New Concord, Mass., 1963, p. 4.)

over a period of two hours. The results, shown in Fig. 15.10, show that reaction time to a faint visual signal under noise increased gradually during the two-hour period.

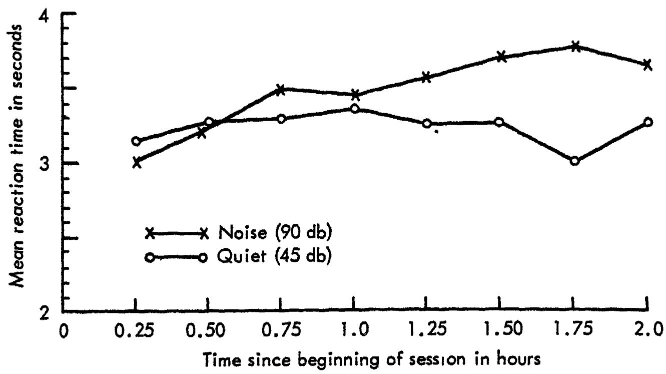


Fig. 15.10. Reaction time to faint visual signals under noise (90 decibels) and quiet (45 decibels). (Based on data from Broussard, *op. cit.*, as plotted by Broadbent, *op. cit.*)

Effects of Noise on Hearing The effects of noise on hearing are much less controversial than are the effects of noise on work performance. Numerous studies and surveys have confirmed the fact that long exposure to high levels of noise brings about some degree of hearing loss.

Typical of various studies of hearing loss is one reported by Glorig and Wheeler[23] of men working in a plant with a noise level of about 105 decibels. Fig. 15.11 shows the average hearing loss of men with various years of exposure to the environment. Each line shows, for the group in question, the average *minimum* intensities of sounds of the various frequencies that could just barely be heard. Fig. 15.11 shows that the greater the length of exposure, the greater the degree of hearing loss, especially in the 4,000 to 8,000 cps range. It should be pointed out that these hearing losses were found to be *greater* than would typically be expected on the basis of age alone. The presumption, then, is that the hearing loss was caused by the plant noise level.

Although precise information is lacking regarding the specific hearing loss that would be brought about by any given noise condition, there is general support for saying that the amount of hearing loss is affected both by intensity of the noise, the duration of the exposure to the noise, the continuity of the noise, the frequency content of the noise, and the

[23] A. Glorig and D. E. Wheeler, "An Introduction to the Industrial Noise Problem," *Illinois Medical Journal,* 107 (No. 1, January, 1955).

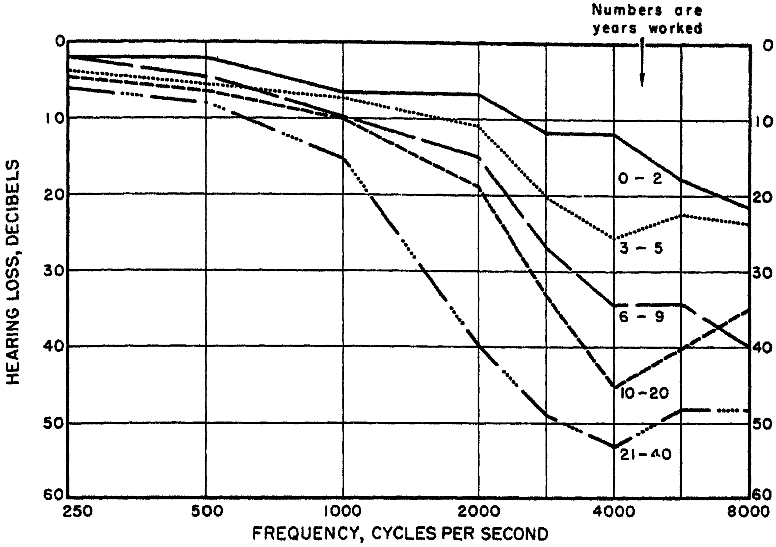


Fig. 15.11. Average hearing loss of men working, for years shown, in noise environment of about 105 decibels. (From Glorig and Wheeler, *op. cit.*)

health, age, and susceptibility of the individual.[24] While these factors cannot be separated completely, it is possible to establish a noise level criterion which will conserve the hearing of a large majority of individuals during relatively continuous exposure to broad-band noise over long periods of time—up to normal working lifetime. This criterion, presented by Williams,[25] is shown in Fig. 15.12. This shows the upper limits that are acceptable for each of various durations of exposure, ranging from single exposure to full-time exposure.

Noise Control The objective of a noise control program in industrial and business organizations is not the elimination of noise, but rather its reduction to reasonable levels, consistent with the type of activity being carried out, and with employee welfare and performance. For extremely noisy circumstances, noise control preferably should be directed *at least* to the reduction of noise to levels that will not cause appreciable hearing loss. Granting that the "safe" level is not highly precise, nonetheless we have seen that noise in the 80 to 90 db level is not generally considered to be "safe."

Noise control can be achieved by various methods, frequently used in combination. These include reduction at the source (such as by proper machine design, proper maintenance, lubrication, mounting equipment

---

[24] L. J. Williams, "Some Industrial Noise Problems and Their Solution," *Noise Control,* 5 (No. 1, January, 1959), 36–43f.
[25] *Ibid.*

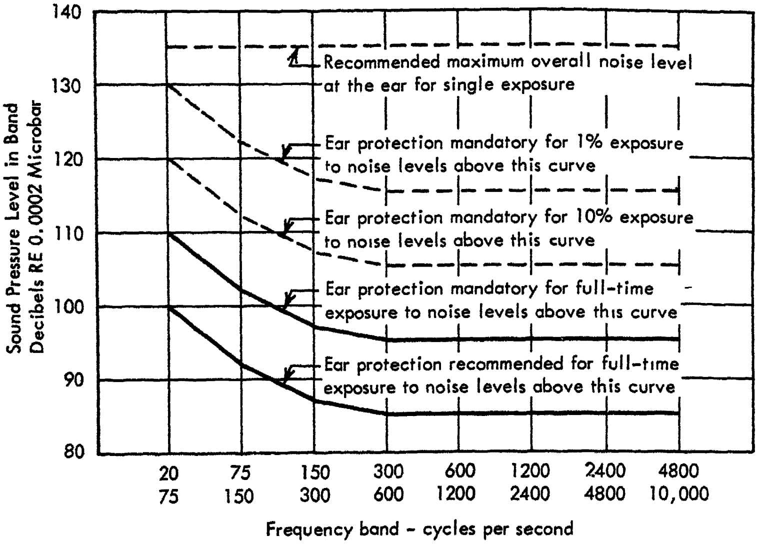


Fig. 15.12. Broad-band noise level criteria for hearing protection. For each indicated duration of exposure, the noise spectrum should not exceed the intensities of the curve in question. (From Williams, *op. cit.*, Fig. 3.)

on rubber); enclosing the noise; use of baffles; use of sound-proofing materials (walls, ceilings, floors, and so on); and the use of ear protection devices such as ear plugs and ear muffs.

MUSIC

During the 1940's and the early part of the 1950's there was quite a flurry of research interest relating to the effects of music during work. In more recent years there has been relatively little research in this area, but there are, of course, many organizations that provide music for their employees. Such music is played for employees on the hypothesis that it is a "good thing" in one way or another—that there are benefits that accrue from such a program, such as improved performance and higher morale. It is, therefore, appropriate to raise the question as to whether music has the beneficial effects that have been postulated for it.

We will not attempt a review of the results of surveys relating to this topic; this has been very thoroughly and effectively done by Uhrbrock.[26] Rather, we will list some of the points that he made in his summary:

[26] R. S. Uhrbrock, "Music on the Job: Its Influence on Worker Morale and Production," *Personnel Psychology*, 14 (1961), 9–38.

1. Unqualified claims that increased production results from the introduction of music into the work situation are not proven.
2. The social implications of music in industry as an incentive system ultimately should be faced. A question may be asked, "Is this a legitimate device that gives pleasure to workers and profit to employers?"
3. Feelings of euphoria during periods of music stimulation have a physiological basis which is evidenced by changes in blood pressure that occur in some subjects while listening to music.
4. Factory employees prefer working where music is played rather than where it is not played.
5. Not all workers like music while they work. From one to ten per cent are annoyed by it.
6. Quality of work can be adversely affected by the use of music in the work situation.
7. Instrumental, rather than vocal, music is preferred during working hours by the majority of workers.
8. There is a negative correlation between age and preference for work music.
9. At least three investigators have reported that young, inexperienced employees, engaged in doing *simple,* repetitive, monotonous tasks, increased their output when stimulated by music.
10. Evidence has been presented which demonstrates that experienced factory operators, whose work patterns were stabilized and who were performing *complex* tasks, did not increase their production when music was played while they worked.
11. At times music has had an adverse effect on the output of individual employees, even though they reported that music was "quite pleasant."

It is apparent from Uhrbrock's summary that music has not proved to be a universal boon to industry, and it should not be viewed as a panacea for all of the production ills of an organization. Perhaps the most relevant considerations in contemplating a music program would be the reactions of employees to it, and the type of work in question. In this latter connection, music probably can be of greater value as a morale booster in the case of employees on simple, routine tasks than on more complex tasks.

HOURS OF WORK

During the last several decades there has been a trend toward the reduction of the work week. With increased mechanization and automation, this trend is likely to continue. But, while the present work week is undoubtedly a very reasonable one for people on many jobs, there are

at least some kinds of work situations in which hours of work are still of some possible concern. One might wonder, for example, about the economy of substantial overtime work, the possible effects of long periods of motor vehicle driving, or the hours of work of people who are in adverse working conditions. In connection with such questions, there is, unfortunately, relatively little relevant research data available. This paucity probaby is in part the consequence of the difficulty of obtaining data from real-life work situations that provide comparisons of various work schedules.

Hours of Work Week in Manufacturing Industries For what interest it might be, a survey was carried out by Kossoris, *et al.*[27] after the war that dealt with the experiences of 34 plants in connection with various work schedules that they used during the war. The survey included 78 "cases" in these plants, covering 2,445 men and 1,060 women. Each "case" that was selected for the survey dealt with a particular organization unit (such as a department or section) for which there had been some change in hours of work from a prewar to war condition or from a war condition to a postwar condition, but for which there were no other major changes. For each case, data were obtained on "efficiency" (i.e., productivity per hour), injuries, and absenteeism.

The data from the survey indicated considerable variation in the effects of changes, depending on how heavy the work was, whether the work was machine paced or not, whether hours were being increased or reduced, method of pay (incentive or hourly), etc. Generally speaking, however, the results indicated that, everything else being equal, the 8-hour day and 40-hour week are best in terms of efficiency and absenteeism, and that higher levels of hours are less satisfactory. With few exceptions, the longer hours resulted in greater output than that produced during shorter schedules. As a rule, however, the increase fell considerably short of the increase in hours. For hours above 8 per day and 48 per week, it usually took 3 hours of work to produce 2 additional hours of output when work was light. When the work was heavy, it took more than 2 more hours to produce 1 hour of additional output. Injuries also increased as hours increased, not only in absolute numbers, but also in the rate of increase.

Efficiency of Continuous Mental Work Within a work period of several hours, there frequently are changes (usually a decrement) in the level of work performance. A decrement frequently occurs, for example, in the case of continuous mental activities. In this connection, Ray, Martin, and Alluisi[28] have presented, in one figure (Fig. 15.13), the

---

[27] M. D. Kossoris, R. F. Kohler *et al., Hours of Work and Output,* U.S. Department of Labor, Bureau of Labor Statistics Bulletin, No. 917, 1947.

[28] J. T. Ray, O. E. Martin, Jr., and E. A. Alluisi, *Human Performance as a Function of the Work-Rest Cycle,* National Academy of Sciences, National Research Council, Publication 882, 1961.

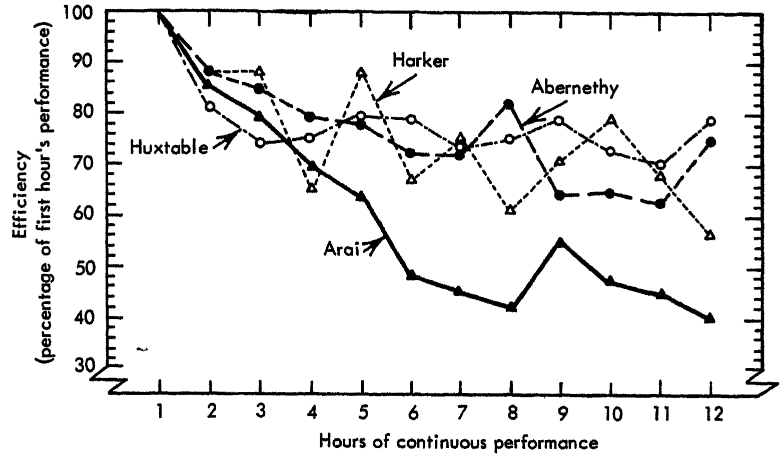


Fig. 15.13. Changes in performance on a mental multiplication task during 12 hours of continuous work. (Data based on studies by Arai, *op. cit.*, and Huxtable, *et al.*, *op. cit.*, and others as summarized by Ray, Martin, and Alluisi, *op. cit.*)

results of the classic study by Arai [29] and the subsequent replication by Huxtable *et al.*[30] While some caution is required in extrapolating from laboratory studies to actual jobs, there are other inklings to substantiate these implications.

Continuous Vehicle Driving As another example of an activity performed over a period of time, let us take the case of vehicle driving. It is, of course, difficult to study driver behavior in relation to driving time under everyday driving conditions. In a controlled experiment Herbert,[31] and Herbert and Jaynes,[32] had 180 male drivers serve as subjects. Five different men served as subjects each of 36 days, one of the five being assigned to one of five experimental conditions, namely, driving for 1, 3, 7, or 9 hours, or serving as a control (with no driving). Before and after completion of the specified number of driving hours, the men were given 9 different driving tests, each test measuring some particular aspect of driving behavior, such as parking, "jockeying," driving over specific courses, etc. The control subjects were given the driving tests without any period of driving on the day in question.

[29] T. Arai, "Mental Fatigue," *Teachers College Contribution to Education,* No. 54 (1912).

[30] Z. L. Huxtable, M. H. White, and M. A. McCartor, "A Re-performance and Re-interpretation of the Arai Experiment in Mental Fatigue with Three Subjects," *Psychological Monographs,* 59 (1946), No. 275.

[31] M. J. Herbert, "Analysis of a Complex Skill: Vehicle Driving," *Human Factors,* 5 (No. 5, August, 1963), 363–372.

[32] M. J. Herbert and W. E. Jaynes, "Performance Decrement in Vehicle Driving," *Journal of Engineering Psychology,* 3 (1964), 1–8.

Without explaining the details, it was found that all of the tests but one were significantly related to driving time. Figure 15.14 shows, for a combination of four of the tests, the decrement in test performance as a function of driving time. While these findings are not necessarily indicative of deterioration in driving behavior or accident rates in actual driving, there is a strong implication of loss in performance over hours of driving.

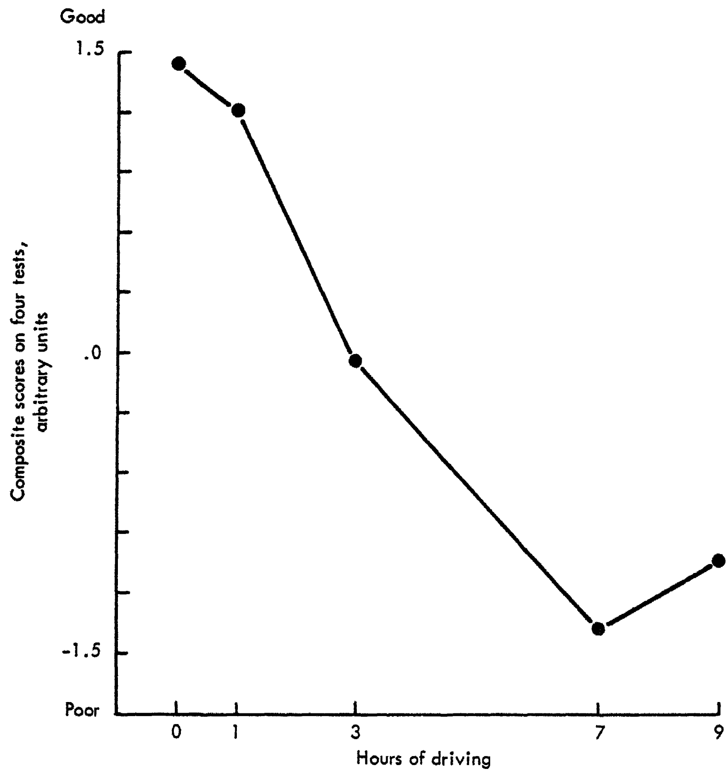


Fig. 15.14. Changes in performance on a combination of four driving tests as related to hours of vehicle driving (The scores are z-scores.) (From Herbert and Jaynes, *op. cit.*)

REST PAUSES

Rest periods are becoming fairly standard practice in many work situations. In some instances such breaks are scheduled, whereas in other instances they are taken at the convenience of the individual. By and

large, both available research and experience support the notion that they have some advantage, not only for the individual, but also in terms of work performance. While some break in the routine of work is probably useful in almost all types of work, there are two classes of work in which breaks are of particular value. One of these is in the case of monitoring tasks in which the attention demands are high. Since Chapter 17 deals with monitoring activities, we will not discuss this type of work further, except to state the fact that performance on such tasks tends to deteriorate after 20 to 30 minutes. These findings suggest that in the case of monitoring activities rest periods, or change in work, should be provided at 20- or 30-minute intervals. The other type is in the case of continued physical work, especially where it is fairly heavy work. In such cases rest pauses are required to prevent excessive physiological fatigue.

An example of the effects of rest pauses in a typical production job comes from an early study in Great Britain by Farmer and Bevington.[33] Figure 15.15 shows the production of a group of employees before and

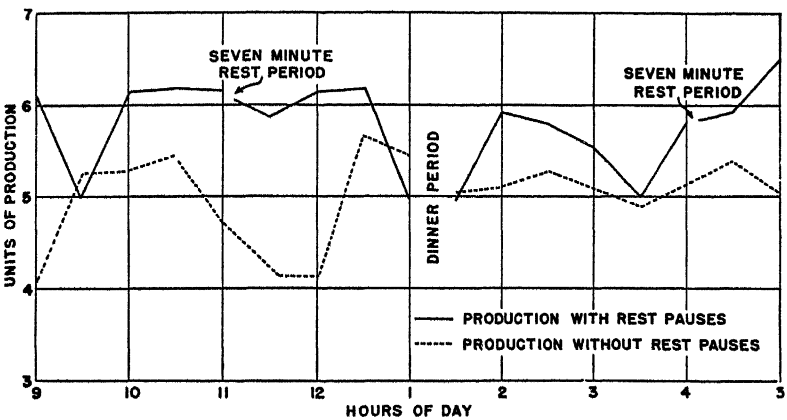


Fig. 15.15. Effect of rest pauses on production for a typical industrial job.

after the introduction of scheduled rest periods. In another study, Miles and Skilbeck[34] found that two 15-minute change-of-work periods resulted in a 14.2 per cent increase in performance.

In considering rest pauses, there is a question as to whether they should be scheduled, or whether they should be taken at the discretion

[33] E. Farmer and S. M. Bevington, "An Experiment in the Introduction of Rest Pauses," *Journal of the National Institute of Industrial Psychology,* 1 (1922), 89–92.

[34] G. H. Miles and O. Skilbeck, "An Experiment on Change of Work," *Occupational Psychology,* 18 (1944), 192–195.

of the employee. In this connection, McGehee and Owen [35] found that the introduction of two short rest pauses during the day reduced unauthorized rest pauses and increased the speed of work. Such findings, however, should not lead to the conclusion that scheduled rest periods are universally preferable. (There are, however, certain administrative considerations that in some cases will argue for scheduling rest breaks. This would be especially the case where a work process would be interrupted by the absence of one individual. In such cases, scheduling becomes almost mandatory. Perhaps especially in heavy physical work or work carried out in hot environments, rest periods usually should be fairly frequent, and employees should feel free to "take five" if they start to feel fatigued. If such voluntary breaks are found to be inadequate to prevent the onset of physiological fatigue, frequent rest periods probably should be scheduled. In such work it is important to prevent a buildup of fatigue, since if a person becomes physiologically impaired, the rest required for recovery is disproportional to the state of fatigue. Frequent short breaks are more effective for recovery than an equal (total) rest period taken at a single time.

There is some evidence that some physical work is performed in such a way that individual muscles or muscle groups get very short breaks in the course of work. In a study by Schmidtke,[36] for example, in which film analyses were made of body movements on nine jobs, it was found that about 18.6 per cent of the time was taken in some form of rest, of the fingers, the arm, or of the total body; about 13 per cent of the time consisted of short breaks in work of a fraction of a minute, these generally being rest pauses of the fingers or arm.

When rest periods are to be scheduled, they usually should be introduced at those times when employees are experiencing the greatest fatigue. An investigation similar to one reported by Griffith, Kerr, Mayo, and Topal [37] is often helpful in identifying the proper times. These investigators found that 379 employees in one plant, surveyed by means of an anonymous questionnaire, rather generally reported greatest feelings of "fatigue" during the fourth and eighth hours of the eight-hour shift. The results were remarkably similar for the manual, office, and supervisory employees covered in this study.

However, as a general practice it is desirable to provide breaks before the buildup of a substantial level of fatigue; this is especially pertinent

---

[35] W. McGehee and E. B. Owen, "Authorized and Unauthorized Rest Pauses in Clerical Work," *Journal of Applied Psychology*, 24 (1940), 605–614.

[36] H. Schmidtke, "Analyse des arbeitsablaufes unter besonderer berucksichtigung der arbeitspausen," *Zeitschrift für Experimentelle und Angewandte Psychologie*, 6 (1959), 248–271.

[37] J. W. Griffith, W. A. Kerr, T. B. Mayo, Jr., and J. R. Topal, "Changes in Subjective Fatigue and Readiness for Work During the Eight-hour Shift," *Journal of Applied Psychology*, 34 (1950), 163–166.

in the case of strenuous physical work. In the scheduling of rest periods it is necessary to determine their frequency, duration, and spacing. Since these factors will naturally vary with the nature of the work, it is impractical to establish any general pattern that would be optimum for a variety of jobs. Ideally, it is, of course, desirable to determine the best rest-period mix for each job separately, by experimentation. Since this is usually not practicable, however, the next best alternative is to adopt, for a given type of job, a pattern of rest periods that has been found through experience to be satisfactory with a somewhat similar type of job. In practice, the duration of rest periods that have been proved beneficial in typical circumstances has varied from two to fifteen minutes, and the number of rest periods during the day has varied from one to five or more (frequently one in the morning and one in the afternoon).

Discussion This chapter has covered briefly some of the more common aspects of working conditions. In certain types of work, however, other aspects of the working environment may be of major importance. These include such aspects as vibration, motion, acceleration, high altitudes, compression chambers (as in underwater construction), and atmospheric impurities. Some of these are discussed in other sources.[38]

---

[38] E. J. McCormick, *Human Factors Engineering* (New York: McGraw-Hill Book Company, 1964).

16

Equipment and Work Design

The nature of jobs is, of course, an important source of variance in work performance. The activities that comprise a job are the consequence of two primary factors: (1) the design of any physical equipment and physical facilities used, and (2) the job methods that have been established by the organization or that are developed by the worker. Since these factors are amenable to change, they can be modified in order to optimize relevant criteria such as productivity, quality, physiological costs, or employee safety and welfare. There are two general domains of effort that are concerned with work activities: methods analysis and human factors engineering.

METHODS ANALYSIS

Historically, the industrial engineers have been concerned with developing work processes, with particular reference to the work activities of people, including work procedures, work layout, work standards, etc. Various techniques of methods analysis are used in developing work

methods and layout.[1] The purposes of methods analysis (which Barnes refers to as motion and time study) are given by Barnes [2] as follows:

1. Developing the preferred system and method.
2. Standardizing this system and method.
3. Determining the time required by a qualified and properly trained person working at a normal pace to do a specific task or operation.
4. Assisting in training the worker in the preferred method.

From ordinary observations and the application of simple logic, it is apparent that different methods of doing a certain job may require different amounts of time and physical energy. The possibilities of improvement by such methods were dramatically demonstrated early in the century by Gilbreth [3] with the operation of bricklaying. He showed that the work of the average bricklayer can be increased from 120 bricks per hour to 350 per hour by following a more efficient pattern of movement. Early studies such as this pointed the way toward the development of the various techniques of methods analysis as they are used by present-day industrial engineers, and of the principles that are applied in the development of work methods.

The principles of motion economy as presented by Barnes [4] are classified in three broad groups, as related to:

1. The use of the human body.
2. The arrangement of the work place.
3. The design of tools and equipment.

Those principles relating to the use of the human body will be presented later. The design of tools and equipment is relevant to human factors engineering.

HUMAN FACTORS ENGINEERING

The second type of effort is that which has come to be known as human factors engineering, human engineering, and (in Great Britain) ergonomics.[5] This field intermeshes with methods analysis, but it is con-

---

[1] Readers who are interested in further aspects of this are referred to such sources as:

R M. Barnes, *Motion and Time Study*, 5th ed. (New York. John Wiley & Sons, Inc., 1963).

G. Nadler, *Work Design* (Homewood, Ill.: Richard D. Irwin, Inc., 1963).

[2] Barnes, *op cit.*, p. 4.

[3] F. Gilbreth, *Bricklaying System* (New York: M. C. Clark, 1909).

[4] Barnes, *op cit.*, p. 220.

[5] Readers who are interested in a further treatment of this topic are referred to such sources as·

E. J. McCormick, *Human Factors Engineering* (New York: McGraw-Hill Book Company, 1964).

R. M. Gagné *et al.*, *Psychological Principles in System Development* (New York: Holt, Rinehart & Winston, Inc., 1962).

C. T Morgan, J S Cook, III, A. Chapanis, and M. W. Lund, *Human Engineering Guide to Equipment Design* (New York: McGraw-Hill Book Company, 1963).

cerned more with the design of equipment and other physical facilities in terms of human considerations. Human factors engineering can be defined as designing for human use. Man has always striven toward designing the things he uses in order that he can use them more effectively and with minimum effort on his part. Since the types of equipment and facilities used in the days gone by were relatively uncomplicated, it was possible to try things out, and if they were not fully suitable for human use, the next version could be modified. Thus, many items of human use (hand tools, etc.) went through an evolutionary process.

In more recent years, however, more systematic attention has been given to the design of equipment and other physical facilities from the human factors point of view. This recent concern was instigated primarily by the experience of the military services in the use of new types of weapon systems. It was found during the war, for example, that some new systems could not be operated effectively because of design deficiencies related to their use by people. The subsequent development of space programs, and of more complicated industrial and other systems, has focused attention on the need to take human factors into account early in the design of such systems, in order to have greater assurance that the systems can be used more effectively when they are produced. This shift has placed greater emphasis on the use of research as the basis for designing items for human use, since the slow, evolutionary process, predicated on human trial and error, is incompatible with today's fast-moving technology. The research that provides the base for human factors engineering comes from such fields as psychology, anthropology, physiology, and biology. The gamut of such research is broad, including such varied topics as visual performance, speech communications, decision-making, performance under weightlessness, and psychomotor performance.

HUMAN WORK IN SYSTEMS

The conventional frame of reference in dealing with personnel functions is that of accepting jobs as "given," and of selecting, placing, and training people for such jobs. On the other hand, at least the earlier focus of those concerned with human factors engineering and methods analysis has been that of so simplifying the tasks to be done that they could be performed by many or most people. A new, and gradually developing, frame of reference is sometimes referred to as the "systems" or "man-machine systems" approach. A major objective of this approach is that of developing systems in which the *combination* of men and machine components and methods is, as nearly as possible, optimum—in which machines perform those functions which they can do best, and men perform those functions which best utilize human talents. At pres-

ent, this objective needs to be viewed as unfinished business, since we do not yet know very well how to go about this "optimizing" process.

Human Functions in Systems In a very general sense, human beings perform three types of functions in the systems of which they are a part. These are: (1) information-receiving; (2) information-processing and decision-making; and (3) action. Consider, for example, an electric power-station operator. He receives information from the instruments of his control panel. On the basis of this information he decides what to do, and he takes an action such as closing one switch and opening another. The basic relationship of men to their work situations is shown in Fig. 16.1. Although work situations naturally vary a great deal, these three basic types of functions are performed on every job.

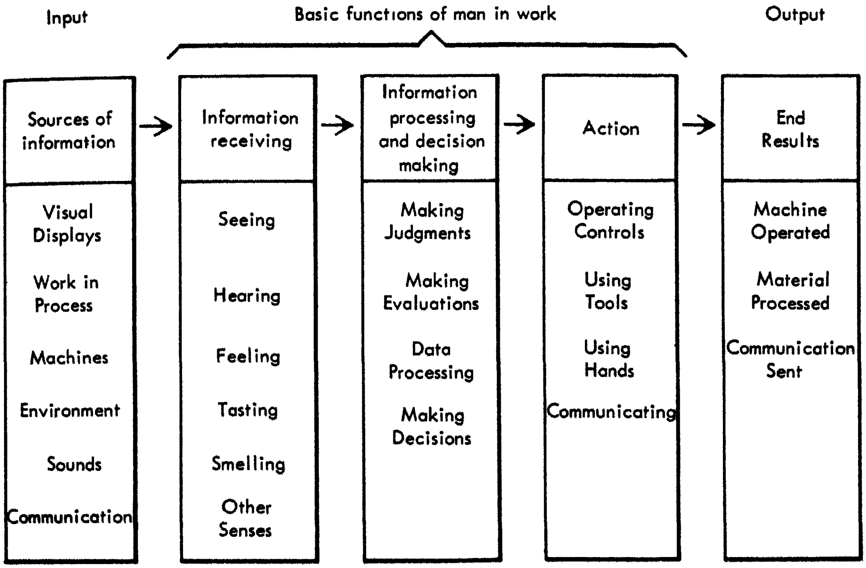


Fig. 16.1. Schematic diagram of basic functions of human beings in man-machine systems. For each basic function a few examples are given in the lower (unshaded) part of the figure.

Efforts to provide for the effective use of human beings in systems, then, need to be directed toward facilitating the information-receiving, the information-processing and decision-making, and the action functions that people are to perform. These functions can be facilitated by the proper design of equipment and other physical facilities, and by the development of appropriate methods and procedures of work. Since it is impractical to deal extensively with these topics, we will summarize briefly a few illustrative examples of relevant research, and will present

some principles of motion economy. This discussion will deal separately with the three major functions mentioned above.

INFORMATION-RECEIVING PROCESSES

The sensory organs of the body are the avenues through which an individual receives information regarding the world about him, including the information that is available to him in performing his job. We commonly think of there being five senses, namely, seeing, hearing, touch, taste, and smell. There are, however, certain other senses, such as sensations of heat and cold, of body movement, of body posture, and of position of body members (and probably some others not yet identified). Seeing and hearing are the senses used most frequently in work situations, although some of the other senses are important on certain kinds of jobs.

In designing the equipment for a job there are, in general, two major considerations in determining what sense or senses to use in the transmission of information to the worker. In the first place, in some circumstances the appropriate sense is, for practical purposes, suggested by the nature of the source of information or by the circumstances under which it is to be received. The fact of a telephone call coming in, for example, can most appropriately be made known to the household by hearing, because of the attention-getting power of sound throughout a house. We might also cite the instance of an electric meter in the home. To convey information about electric-power consumption in the home, one would certainly not want a noise stimulus; thus, a meter provides a record that can be seen visually These two examples are clear cut; but there are many other situations in which the appropriate sense avenue is not nearly as straightforward. For example, although information about the speed of the family car typically has been presented visually, one could argue for the advantage of having a sound signal to indicate an appropriate maximum speed.

In the second place, where either of two (or more) sense avenues could be used in the transmission of information, the determination of the method to be used may be based on consideration of the relative demands already imposed on the various sensory avenues. Where, for example, the visual sense is already burdened, perhaps the auditory or the tactual senses might be brought into use to relieve the visual sense. This would be particularly desirable in complex operations where a great deal of information has to be received and integrated, such as in the operation of aircraft and some types of military equipment.

In considering various possible sensory avenues for use in information transmission, one might wonder about possible differences in reaction time to different stimuli. Fig. 16.2 shows typical reaction times to various types of stimuli. It is relevant to note that the differences in reaction time for hearing, touch, and sight are relatively negligible; these

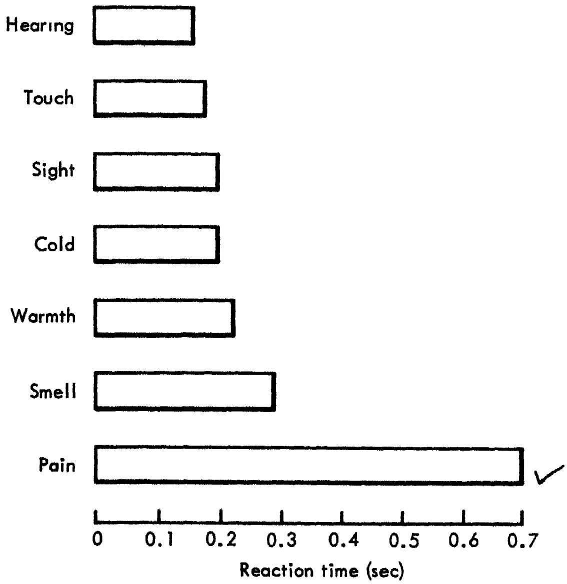


Fig. 16.2. Reaction time to stimuli presented through various senses. (From Morgan, *et al., op. cit*, p. 229.)

are the senses that are used most extensively in man-machine systems. It should be noted, however, that reaction times are influenced by numerous other factors, such as the specific nature of the stimulus, the intensity and duration of the stimulus, and the number and type of competing stimuli, as well as by the age and sex of the individual. (Reaction time usually is shortest for individuals in their twenties.)

Visual Displays The term *visual display* applies to any device that is used to present information visually to human beings. Examples include dials and other visual instruments, signal lights, signs, and so forth. A great deal of the research in human factors engineering to date has been directed toward the design of visual displays that convey the information most adequately to human beings.

Instrument design. Visual instruments are used for one of three purposes, as follows:

1. Quantitative reading (to determine a quantitative value, such as a temperature reading).
2. Qualitative reading (to identify one of three or more "conditions," such as *high, average,* or *low*).
3. Dichotomous reading or check reading (to identify one of two conditions, such as *on* or *off,* or *normal* or *not normal*).

Some examples of instruments are shown in Fig. 16.3. These are grouped into moving pointer (fixed scale) and moving scale (fixed

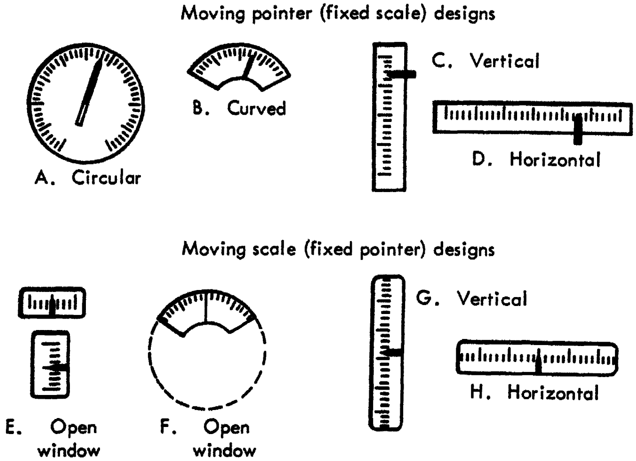


Fig. 16.3. Examples of moving pointer and moving scale designs of visual instruments.

pointer) classes. There have been various studies relating to the speed and accuracy of reading instruments. In an experiment by Elkin [6] the circular (A), vertical (C), and open-window (E) designs were used in both quantitative and qualitative reading tasks. In different phases of the study, three different time limits were used, as well as a no-time limit condition. In the quantitative reading task, the open window was read with shortest time and fewest errors, followed by the circular and vertical in that order. But the differences were greatest for the 0.12-second time limit; for longer time exposures (0.36 and 1.08 sec.) and no-time-limit, however, the differences were much less accentuated and generally not significant. In the qualitative reading phase of the study, the order of superiority of the three scales was different, the order being circular, vertical, and open window.

These findings (and those from other studies) suggest that the best design of an instrument varies with its purpose. The open-window design (or a counter), for example, is generally satisfactory for quantitative reading of values that are not subject to frequent or rapid change. While the vertical and horizontal scales (C and D) are useful for short scales, they are not as satisfactory for noticing changes as, say, the circular form (A). The moving scale designs, in turn, offer advantages over the moving pointer designs for use with long scales.

Arrangement of dials for check reading. Sometimes many dials

[6] E. H. Elkin, *Effects of Scale Shape, Exposure Time, and Display-Response Complexity on Scale Reading Efficiency*, USAF, WADC Technical Report 58–472, February, 1959.

are arranged together into a bank of dials for check reading (to determine whether each dial is, or is not, at a "normal" condition).

In a study by Senders[7] a comparison was made of the time required for reading groups of dials when the "normal" position in all cases was at the same relative location (namely, at the 9 o'clock position), as opposed to reading groups of dials when the "normal" position was different for each dial. She found that 32 "aligned" dials could be read in the same amount of time (0.5 second) as 4 "nonaligned" dials.

In another study, Johnsgard,[8] in turn, investigated the speed of checking nonaligned pointers of groups of dials arranged in four different ways, as shown in Fig. 16.4. Each subject was presented with 18 sheets

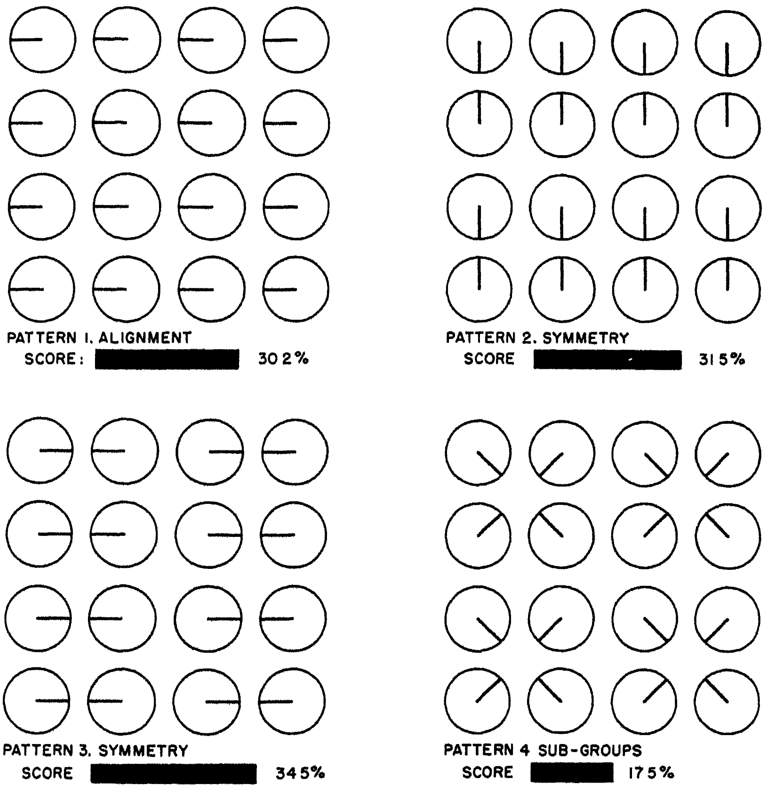


Fig. 16.4. Patterns of dials used in check-reading experiment, showing mean accuracy scores in identifying out-of-normal positions (Adapted from Johnsgard, *op. cit.*)

[7] Virginia L. Senders, *The Effect of Number of Dials on Qualitative Reading of a Multiple Dial Panel*, USAF, Wright Air Development Center, Technical Report 52–182.

[8] J. W. Johnsgard, "Check-Reading as a Function of Pointer Symmetry and Uniform Alignment," *Journal of Applied Psychology*, 37 (1953), 407–411.

with dials arranged in each of the four patterns, making a total of 72 sheets. Some of the sheets had pointer settings that were *not* in the "normal" position, and the subject was to identify such out-of-normal dials. Each subject saw a sheet for half a second.

The mean accuracy scores for the four patterns are shown by the bars underneath each of the four patterns in Fig. 16.4. This shows that the subgroup patterns were read less accurately than the other three patterns. Although the differences among the other three are small, pattern 3 (one of the "symmetry" patterns) was slightly superior to the others.

Auditory Displays While most "displays" used in man-machine systems are visual, in some circumstances auditory signals are more satisfactory for transmitting information. A general set of guidelines for using auditory versus visual presentation is given in Table 16.1.

Table 16.1

GUIDELINES FOR USING AUDITORY VERSUS VISUAL MODES OF PRESENTATION *

Use auditory presentation if:	*Use visual presentation if:*
1. The message is simple	1. The message is complex
2. The message is short	2. The message is long
3. The message will not be referred to later	3. The message will be referred to later
4. The message deals with events in time	4. The message deals with location in space
5. The message calls for immediate action	5. The message does not call for immediate action
6. The visual system of the person is overburdened	6. The auditory system of the person is overburdened
7. The receiving location is too bright or dark-adaptation integrity is necessary	7. The receiving location is too noisy
8 The person's job requires him to move about continually	8 The person's job allows him to remain in one position

* *From Morgan* et al., op cit, *p. 125.*

INFORMATION PROCESSING AND DECISION-MAKING

In performing work activities, some information processing or mediating function occurs between the receiving of information and the taking of some action. The nature of this mediation function naturally varies with the situation, but can include making judgments and evaluations, data processing (such as in computations), reasoning, etc. Whatever this nature, however, the end result is some decision (or choice of action), although this choice may be virtually predetermined by the nature of the input.

The distinction between the process of receiving information (which

is, initially, a sensory process) and that of processing such information and making decisions based on such information is difficult to draw with precision. This is because the simple sensory process of receiving a stimulus through a sense organ is inextricably intertwined with *perception*. Perception, as a psychological function, involves the attachment of *meaning* to that which is sensed. Let us take, for example, two people looking at the pressure-gauge of a steam boiler, both of whom are equally able to "see" the reading. One person knows nothing about boilers, so to him the information is meaningless. The other one, who knows about such things, sees that the pressure is very near what he knows to be the danger point The "perceptions" of these two people of the same, identical instrument would be very different. The decision on the part of the second person to reduce the pressure would be almost an automatic one.

The Nature of Decisions to Be Made Given certain informational input, the decision function is that of determining what action would bring about the most appropriate output. In repetitive types of operations (such as putting two parts together) the decision as to what to do is virtually predetermined. In the case of many other types of activities, however, the decisions regarding the actions to be taken must be selected on the basis of many complex inputs and other considerations, such as in serving as a control tower operator at an airport, or operating a bulldozer over rough terrain.

The more "structured" or automatic the relationship between a given informational input and the desired output, the greater is the feasibility of automating the process. Where a specific predetermined action is to be taken for each possible input, the process can be considered as being completely "programmed." While programming typically is thought of as being applied to automated or highly mechanized operations, one can in a sense consider personnel training as a programming process—of training people to make decisions that are appropriate to the information inputs.

With respect to human decision-making abilities, it is generally the case that the decision-making process is more rapid when the *number* of possible (predetermined) choices of action is limited. In this connection Fitts [9] has summarized a number of pertinent studies, including one by Hick.[10] In this study the subjects were asked to respond to appropriate signals by moving fingers that corresponded to the signals. The number of signals (and corresponding number of fingers) varied from one to ten in different parts of the experiment. Thus, if six signals were being used, the subject, upon seeing a particular signal, would have to move the

[9] P. M. Fitts, "The Influence of Response Coding on Performance in Motor Tasks," in *Current Trends in Information Theory* (Pittsburgh: University of Pittsburgh Press, 1953).

[10] W. E. Hick, "On the Rate of Gain of Information," *Quarterly Journal of Experimental Psychology*, 4 (1952), 11–26.

finger that corresponded with that signal. The time required for making such choices is given in Fig. 16.5. These times are usually referred to as *reaction time,* but we can think of this as "decision" time. We can see that the decision time varies with the number of choices from which a selective choice is to be made. The decision time, incidentally, increased logarithmically with increases in number of alternate choices.

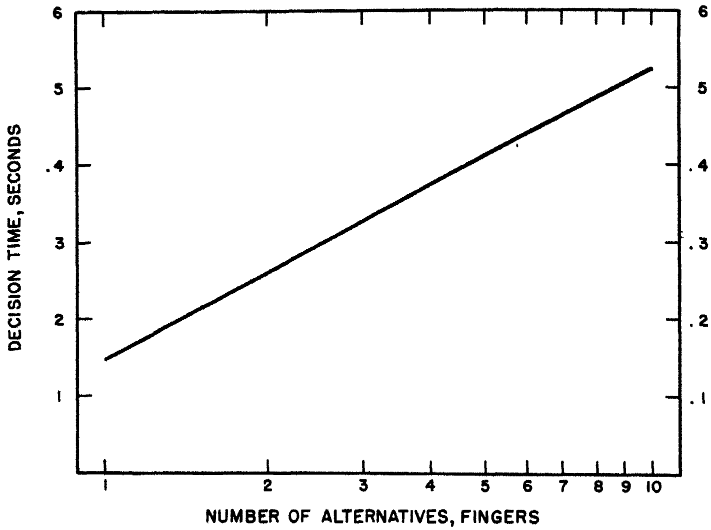


Fig. 16.5. The relationship between decision time and number of possible alternate responses from which to choose in making responses to appropriate signal. (Data from Hick, *op. cit*, as adapted by Fitts, *op. cit.*)

This study probably represents an oversimplification of the general principle in question, but even so it suggests that, where the decisions (choices of responses) are to be made on the basis of presented information, there is a time advantage (and reduced likelihood of error) when the decision is simple rather than difficult, such as having to choose from among a limited number of alternatives than from many. A generalization of this is shown in Fig. 16.6.

But, while human response usually is more rapid when the responses to specified inputs are predetermined and limited in number (i.e., when the decision-making is "programmed"), it is rather clearly evident that one of man's most important abilities is that of making decisions on the basis of complex assortments of information. In the design of systems, therefore, it would seem that, where relevant, the system be so designed as to take advantage of man's superior abilities to make complex decisions and to adapt and improvise in the case of new and varied circumstances.

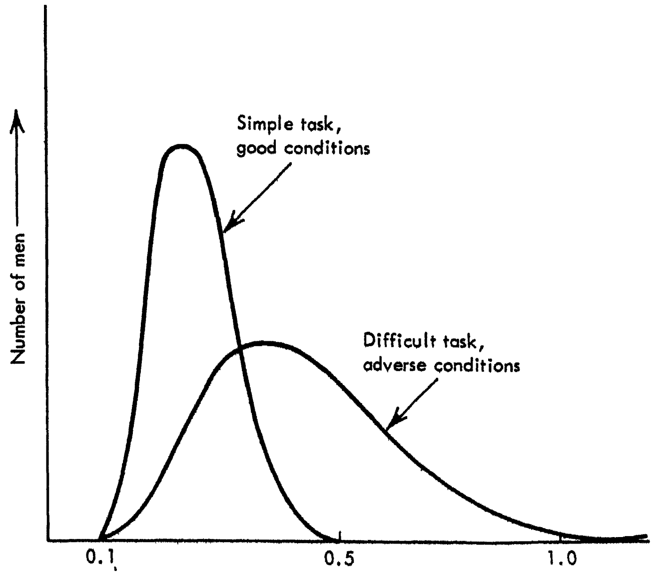


Fig. 16.6. Generalized illustration of differences in decision time (reaction time) in simple decision tasks versus difficult decision tasks. (Adapted from Morgan, *et al., op. cit.*, p. 228.)

Facilitating Decision-Making In the case of certain types of decisions to be made in the operation of man-machine systems, the ready availability of relevant information may have a significant effect upon the adequacy and finality of the resulting decisions. Appropriate information is a requirement for many decisions. This harks back to the matter of information displays. In the design of man-machine systems (such as, say, a control system for a refinery) an analysis preferably should be made of the *decisions* that men would be expected to have to make in actual operation of the system. Knowing what decisions are to be made, it is then in order to determine what *kind(s) of information* would be required in making such decisions. Provision should then be made for providing for such information through the use of appropriate information displays. Information should be provided in such a way that it is most effective in facilitating the decision-making process. While much remains to be learned about this, an example may suggest the type of analysis that can contribute answers to this question.

A study was carried out by Forbes [11] for purposes of studying three

[11] J. W. Forbes, "Auditory Signals for Instrument Flying," *Journal of the Aeronautical Sciences,* **13** (1946), 255–258.

auditory signals used in flying. We will not bother to discuss the procedures of the experiment, or the specific results, but rather we will simply summarize the implications. It was found that a combination three-in-one sound signal was better than three separate signals. The combination signal would be varied in three different ways (specifically in steadiness, frequency, and modulation) to provide information of three different types.

A graphic illustration of the three components of the combination signal is given in Fig. 16.7. The implications of this type of research

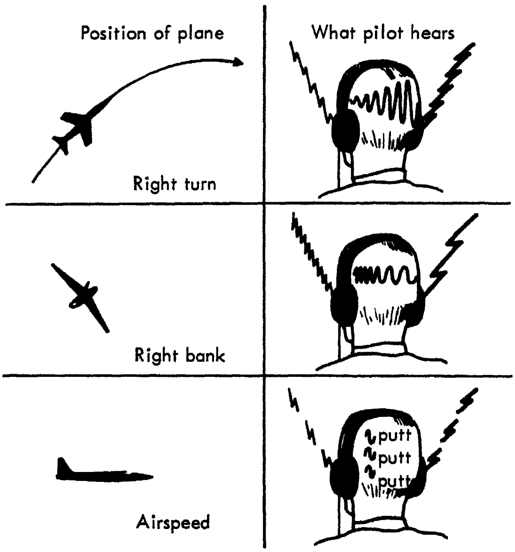


Fig. 16.7. Illustration of three components of a combination signal, the components representing turn, bank, and air speed. (From Forbes, *op. cit.*, as illustrated by Morgan *et al.*, *op. cit.*, p 136)

strongly suggest that the appropriate *integration* of information for workers can facilitate the decision-making process.

ACTION PROCESSES

For a decision to be implemented, some action must be taken by the individual. This may be some physical action or a communication action. For our brief discussion here, we will concern ourselves with physical actions. In the design of equipment and the development of methods it is, of course, desirable to provide for those human physical actions and

responses that will result in the desired output. This can be done through the design, location, and arrangement of control devices, and through the methods that are established for doing the job.

Blind-Positioning Movements Frequently on a job the worker makes what is called a "blind-positioning" movement to some control device such as a lever or push button without looking, or, in some circumstances, without being able to look. The accuracy with which blind-positioning movements can be made in different directions is then a potentially important consideration in deciding where the control device should be located.

A study by Fitts[12] dealt with this. The subjects in the experiment were blindfolded, then seated in the enclosure shown in Fig. 16.8. They

Fig. 16.8. Experimental situation used in study of accuracy of blind-positioning movements carried out by Fitts, *op. cit.* (Courtesy Psychology Branch, Aero-Medical Laboratory.)

were given a sharp-pointed marker and were asked to make blind-positioning movements to the various targets around the enclosure. There were three tiers of targets, the middle one being at approximately shoulder level, and the others being about 45° above and below.

Accuracy was determined by "scoring" the position of the hits made with the marker; a bull's-eye was scored 0, a mark in the first circle around the bull's-eye was scored 1, and marks in the other rings were given

[12] P. M. Fitts, "A Study of Location Discrimination Ability," in *Psychological Research on Equipment Design,* Ed. P. M. Fitts, USAF, Aviation Psychology Program, Research Report No. 19, Superintendent of Documents, 1947.

higher scores up to the outside ring, which was scored 6. Thus, a low score indicates good accuracy. The results are shown in Fig. 16.9. The size of each circle indicates the relative error-scores of the movements to the target at that particular location. The average error-scores are given inside the circles. Fig. 16.9 indicates that accuracy is greatest in the positions straight ahead and is least in the extreme side positions. Considering target level, the accuracy is greatest for the lowest tier and least for the highest tier.

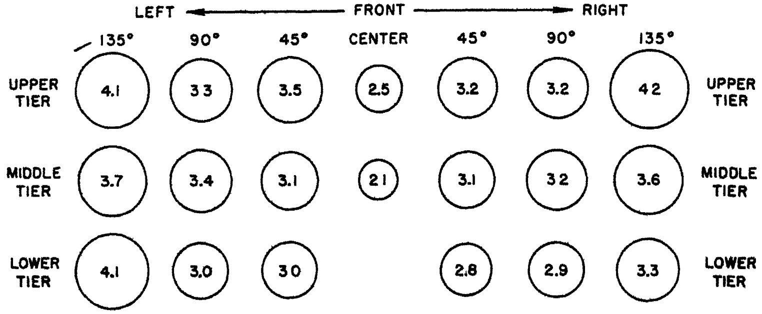


Fig. 16.9. Accuracy of blind-positioning movements in different directions. Each circle represents the location of a target (left to right, and in upper, middle, and lower tiers). The average error-score for each target is given inside the circle, and the size of the circle is proportional to the error-score. (Adapted from Fitts, *op. cit.* Courtesy Psychology Branch, Aero-Medical Laboratory.)

The results of a study such as this suggest that when positioning control devices that are likely to be used without looking, greater accuracy in locating them could be achieved by placing them toward the center (ahead of the worker or slightly to his right or left), and preferably below shoulder height.

Size of Handwheels Many machines or other equipments have handwheels that serve some function such as moving the carriage of the machine to a particular position. Since the purpose and method of use will in large part determine the suitability of various sizes of handwheels, it is obviously not possible to use the same size of handwheel for all purposes. Some, for example, have to be rotated rapidly for many turns; others have to be used to overcome considerable resistance; still others may have to be used with great precision.

To illustrate the type of study that can be carried out in analyzing handwheels for a particular type of purpose, let us cite a portion of a study by Davis.[13] He experimented with various sizes of handwheels which

[13] L. E. Davis, "Human Factors in Design of Manual Machine Controls," *Mechanical Engineering*, 71 (October, 1949), 811–816.

the subjects were to use for the precise positioning of a pointer. Generally speaking the movement required about one revolution. Another variable that he introduced was the amount of torque in the wheels (0, 40, and 90 inch-pounds). He recorded the time required to make settings with different combinations of size and torque, the results being shown in Fig. 16.10. It can be seen that with no torque there is not much difference

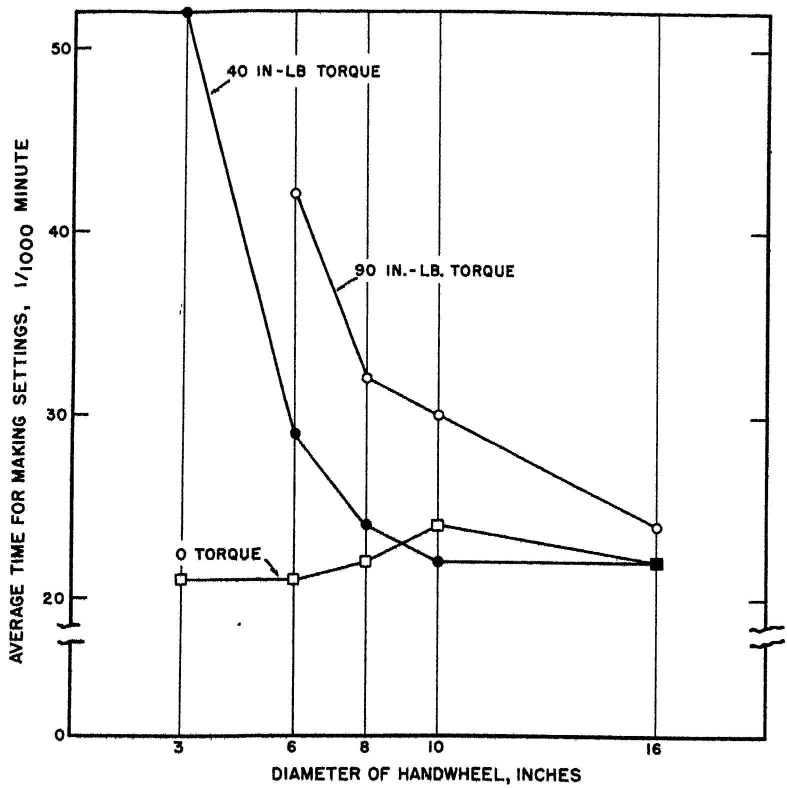


Fig. 16.10. Average times required for making pointer settings with handwheels varying in size and torque. (Adapted from Davis, *op. cit.*)

in time for making a setting among the various sizes, although the three-inch and six-inch handwheels took slightly less time than the larger ones. When using handwheels with substantial torque, however, the size of the handwheel is a very definite factor in the speed of use. Although with a 40 inch-pound torque the 10-inch and the 16-inch handwheels gave equal times, with the 90 inch-pound torque the 16-inch handwheel was distinctly superior to the others.

The results of an analysis of this type would be very helpful to a design engineer in selecting a handwheel that would be most rapidly used (under a given torque combination) for a purpose such as the one used in this study.

Compatibility of Controls Control devices generally should be compatible with human "expectations" in terms of their movement and spatial relationships. In the case of many control devices, for example, the direction of movement of the device has some relevance to the response of the system. Frequently, the response of the system is shown by a visual display. (It should be noted, however, that in some circumstances the control action is to correspond with the display indication, whereas in others the display indication reflects the consequences of the control action.) Where there is some relationship between control movements and corresponding displays, the directional relationships should be compatible (i.e., in the expected direction). Some examples of compatible control-display relationships are given in Fig. 16.11.

Aside from the compatibility of movement relationships, corresponding controls and displays should be so arranged that they have compatible spatial relationships, such as illustrated in Fig. 16.12.

Coding of Control Devices Where a number of control devices of the same general class are to be used, mistakes may occur because of failure to distinguish one from another. Under such circumstances some form of coding usually can reduce such errors. Different methods of coding can be used, such as shape, size, location, and color. Examples of shape coding are shown in Fig. 16.13. Shape-coded controls should be sufficiently different so that they can be identified by feel, which is the case with those illustrated. Those illustrated, however, also have another distinguishing characteristic, namely, the symbolic association of the shape of the control with the use in question. This has the effect of facilitating the learning of the shapes (as related to their uses).

Motion Economy as Related to the Human Body Mention was made earlier of the principles of motion economy as related to the human body.[14] These are relevant to the physical action functions in work, and, therefore, have some implications for the development of work methods and for the design of equipment. Barnes[15] presents nine specific principles. The first three are as follows:

1. The two hands should begin as well as complete their motions at the same time.
2. The two hands should not be idle at the same instant except during rest pauses

---

[14] Barnes, *op. cit.*, p. 222.
[15] *Ibid.*, Chap. 17.

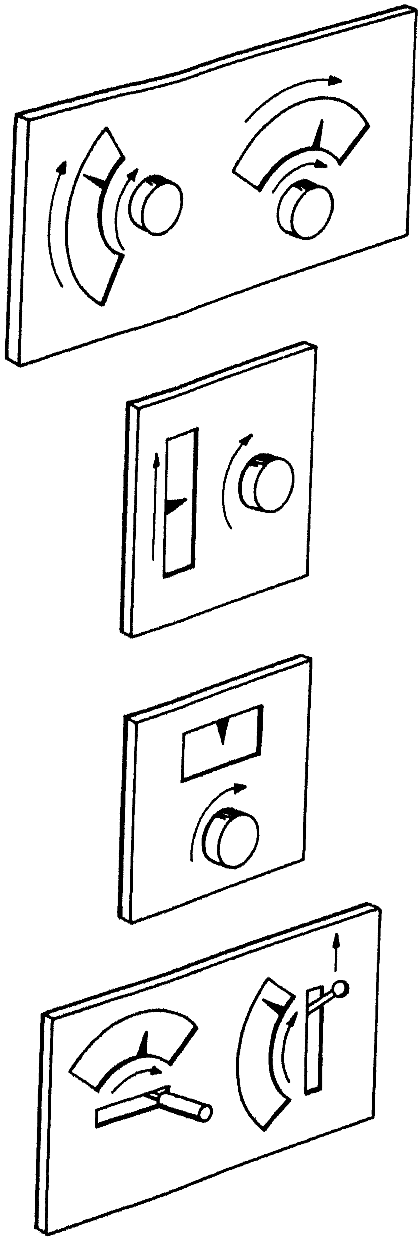


Fig. 16.11. Some examples of compatible movement relationships of controls and displays.

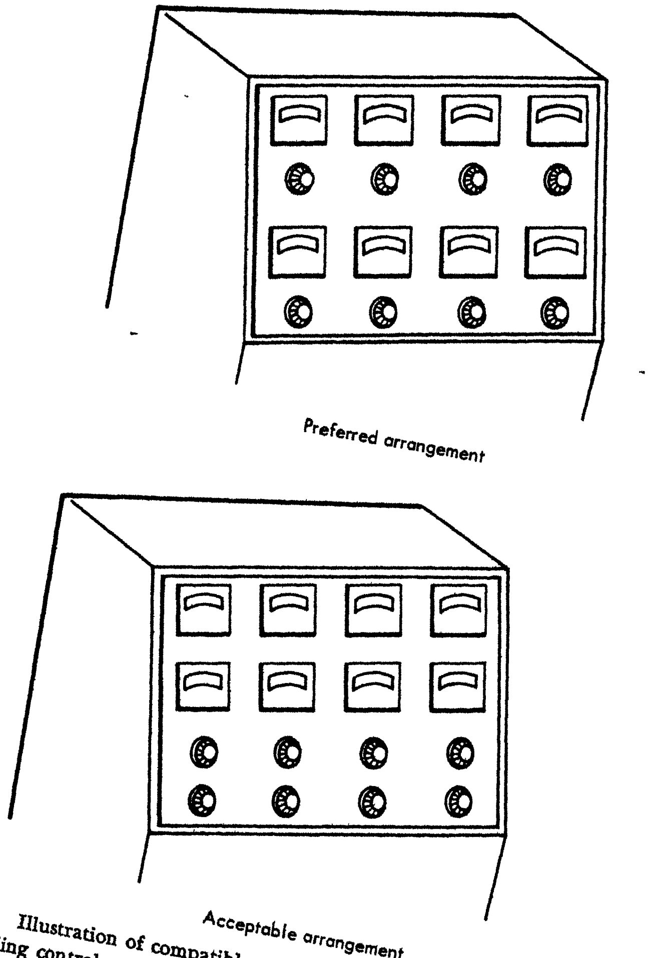


Fig. 16.12. Illustration of compatible spatial arrangements of displays and their corresponding controls.

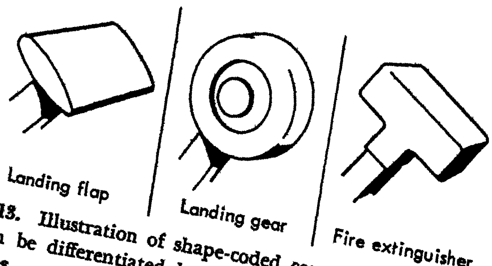


Fig. 16.13. Illustration of shape-coded control knobs. Aside from the fact that they can be differentiated by touch, they also have symbolic association with their uses.

3. Motions of the arms should be made in opposite and symmetrical directions and should be made simultaneously.[16]

Numerous experimental demonstrations of the time saved by following these principles have been made. Several such experiments are reported by Barnes,[17] typical of which is the job of assembling a bolt and three washers. On this job, in the study reported, it was found that a 53 per cent increase in production resulted from a work rearrangement to permit adherence to the three principles listed above.

4. Hand and body motions should be confined to the lowest classification with which it is possible to perform the work satisfactorily.

The five general classes of hand motions include, first, the fingers by themselves, with the others involving, progressively, other body members in this order: wrist, forearm, upper arm, and shoulder. Generally speaking, lower classifications require less time and physical effort than the higher classifications. There are, however, exceptions to this generalization. Controlled investigations have shown, for example, that for certain types of activity, wrist motion and forearm movements are easier, faster, and more uniform than finger motions. Evidence generally shows that the forearm is the most desirable member to use for light work, and that in highly repetitive work, motions about the wrist and elbow are in all respects superior to those of the fingers or shoulders.[18]

5. Momentum should be employed to assist the worker whenever possible, and it should be reduced to a minimum if it must be overcome by muscular effort.

This principle was used in F. B. Gilbreth's early studies of bricklaying.[19] Methods were developed that made use of the momentum of the moving brick in forming the mortar into the joints.

6. Smooth continuous curved motions of the hands are preferable to straight-line motions involving sudden and sharp changes in direction.

One investigation has shown that 15 to 25 per cent of a work cycle is used in changing hand direction when an abrupt change in direction is made.[20] If the work layout can be arranged to eliminate this kind of lost time, a marked saving in time is effected.

7. Ballistic movements are faster, easier, and more accurate than restricted (fixation) or "controlled" movements.

Although there is little experimental evidence on this subject, the experience of industrial engineers indicates that ballistic movements, in com-

---

[16] These three principles were first formulated by F .B. and L. M. Gilbreth, "A Fourth Dimension for Measuring Skill for Obtaining the One Best Way to Do Work," *Society of Industrial Engineers Bulletin*, 5 (1923), 6.

[17] Barnes, *op. cit.*, pp. 223–226.

[18] Barnes, *op. cit.*, p. 237.

[19] F. B. Gilbreth, *Motion Study* (New York: D. Van Nostrand Co., Inc., 1911).

[20] Barnes, *op. cit.*, p. 241.

parison with controlled movements, are more powerful, more accurate, faster, and less fatiguing.[21]

8. Work should be arranged to permit an easy and natural rhythm wherever possible.

Rhythm has been defined in at least two ways—as a regular sequence of uniform motions, and as a regular sequence of accented motions. (An "accented" rhythm is one in which there is an opportunity to accent certain points in the cycle, such as in the sudden thrust of feeding a sheet of material in a punch press.) A rhythm of either type generally tends to facilitate physical work activities, especially in the case of repetitive operations where the operator can do the work with a minimum of mental effort.

9. Eye fixations should be as few and as close together as possible.

Measuring effects of methods Principles of motion economy such as those mentioned above have found wide acceptance throughout industry as a basis for changing (i.e., improving) methods of work. There have been many circumstances in which the methods developed from such principles have been found, through objective study, to be superior in some respect to original methods. Some examples of such studies were mentioned above, in the discussion of certain of the principles listed.

A rather unique study of the effects of a change in method of work is described by Brouha.[22] The work in question involved, in part, heavy lifting and the use of a comparatively heavy hammer above shoulder level. Partial mechanization of the work eliminated these aspects of the job. Heart rate recovery curves were used in comparing the physiological stress before and after the change in method. The "before" and "after" curves are shown in Fig. 16.14. The difference between these indicated very clearly the reduced physiological costs of the new method.

Generally speaking, the evidence from most pertinent investigations tends to support the general principles and procedures of methods improvement. It should be pointed out, however, that the application of a set of methods-improvement principles does not in all cases insure the development of an optimum method.

A study reported by Lauru,[23] for example, raises some question about the general applicability of such principles. This report cites research on work activities with the Lauru platform, which was described in Chapter 15. A comparison was made with the Lauru platform of the

---

[21] *Ibid*, pp. 245–247.

[22] L. Brouha, "Fatigue—Measuring and Reducing It," *Advanced Management*, 19 (January, 1954), 9–19.

[23] L. Lauru, "The Measurement of Fatigue, Part 2," *Manager*, 22 (1954), 369–375.

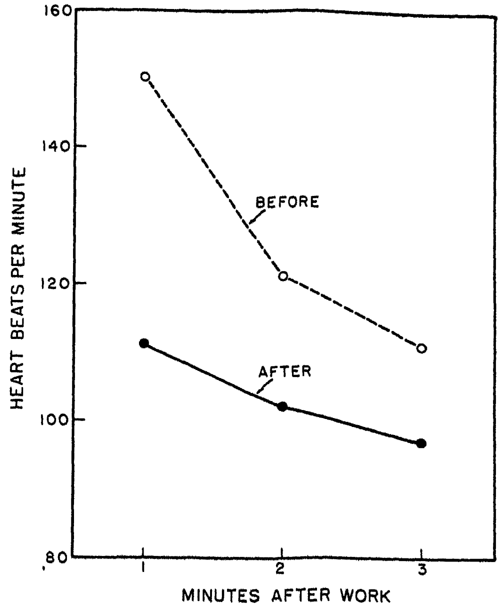


Fig. 16.14. Heart rate recovery curves for manual job before and after partial mechanization. (From Brouha, *op. cit.*)

physical forces created by the movements of a bricklayer working by three different methods, as follows:

1. The conventional method.
2. The "improved" method developed by Gilbreth, mentioned earlier in this chapter; this method involves particularly the principle of bimanual *symmetry* of movements (moving the arms in opposite and symmetrical directions—principle No. 3).
3. A new method involving the use of only one hand at a time (thus departing from the principle of symmetry).

The forces resulting from these three methods are shown in Fig. 16.15. In studying these, it should be kept in mind that the variability of the curves above and below the base line is indicative of the magnitude of the forces created by the worker during the task. It will be seen that there is noticeably less variation with the new method than with the others, including Gilbreth's, thereby suggesting that the new method requires less physical exertion than the others.

It should also be added that the new method, aside from requiring less physical exertion, also took less time; there was a saving of 18.4 per cent in time over the original method, and of 6.4 per cent over the Gilbreth method.

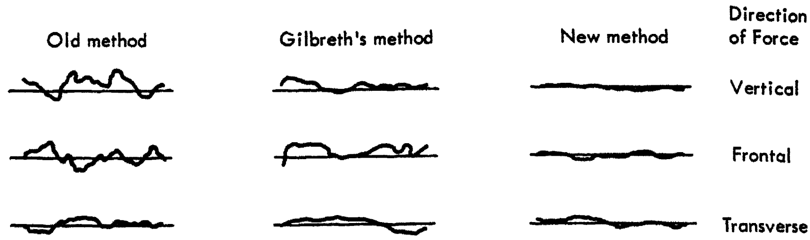


Fig. 16.15. Physical forces exerted when laying brick by three different methods. Forces were measured by use of the Lauru platform in three directions, namely, vertical, frontal, and transverse. The third bricklaying method required the least exertion. (Adapted from Lauru, *op. cit*)

In particular, this study seems to cast some doubt about the general validity of the principle of bimanual symmetry. This should not be taken to suggest that this principle has no applicability in methods improvement, but perhaps it does suggest that further research is needed to determine the work circumstances under which the principle is, and is not, applicable. Perhaps other principles of motion economy also could well be subjected to similar research.

WORKSPACE

An important consideration in the performance of manual work activities is that of the physical space within which the work is carried out, and the arrangement of the various features of it. An illustration of research carried out in this phase of human factors engineering is a study by Squires [24] in which he was concerned with the shape of the "normal work area," such as on a work bench or table. The normal work area traditionally has been considered to consist of overlapping semicircles, each one being formed by the rotation of the forearm with the elbow as a pivot.[25] The study by Squires, however, indicates that the normal work area is somewhat more limited than is traditionally thought. This area is shown in Fig. 16.16.[26] It takes into account the fact that the elbow typically moves out and to the side as the forearm is rotated about it. It also shows the effect of limiting outer position of the forearm-hand as it rotates outward.

This shape is suggested by Squires not only as the normal work-

[24] P. C. Squires, "The Shape of the Normal Work Area," *U.S. Naval Medical Research Laboratory Report No. 275,* 23 (July, 1956).

[25] Barnes, *op. cit.,* p. 260.

[26] For the mathematically inclined, it might be added that the curve is a *prolate epicycloid.*

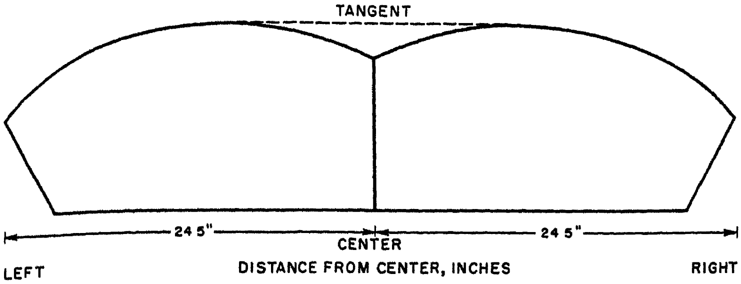


Fig. 16.16. Normal horizontal work area as presented by Squires, *op. cit.* The curves show the areas that fall within convenient reach of the hands when the forearms are rotated around the elbow. The elbow itself tends to move during such rotation.

surface area of a work station, but also as the contour of a vertical panel with control devices. Such a proposed panel is shown in Fig. 16.17. Here the horizontal surface has been made straight, along the tangent of the two curves shown in Fig. 16.16; and the corresponding vertical surface in Fig. 16.17 has been made correspondingly flat.

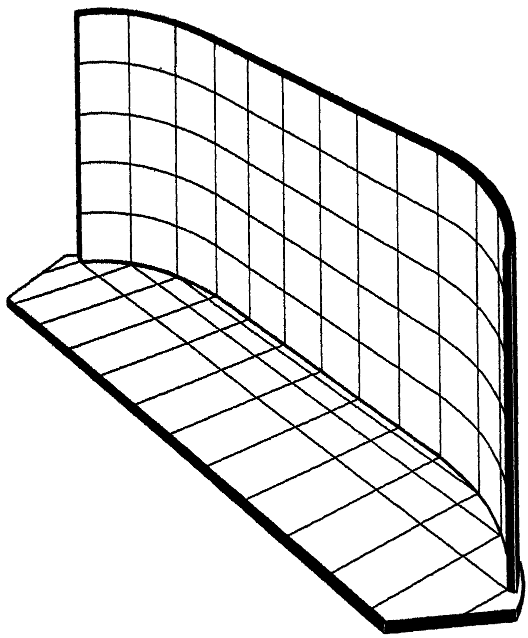


Fig. 16.17. Horizontal and vertical work surface areas as adapted from Fig. 16.16. (Based on study of normal work areas by Squires, *op. cit.*)

The physical dimensions of people, and the range and ease of movements of body members, impose various types of constraints on the arrangement of work facilities. While it is not feasible here to deal very extensively with this aspect of equipment design,[27] Figs. 16.18 and 16.19 will at least suggest some of the types of consideration that are relevant.

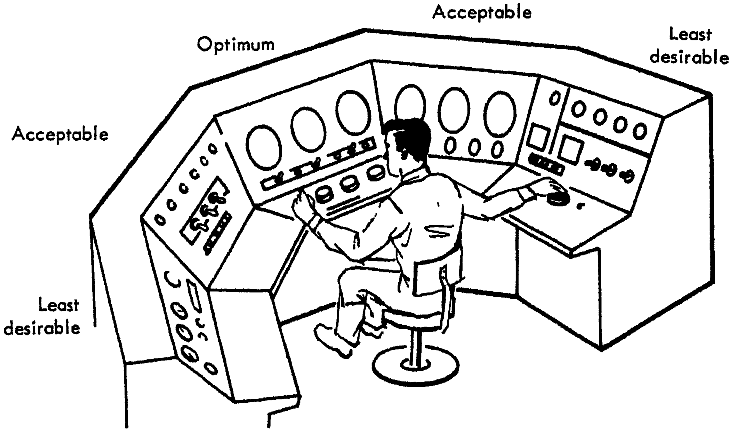


Fig. 16.18. Illustration of work-console arrangement, indicating relative desirability of different sections for displays and controls. (Adapted from Morgan, *et al., op. cit.*, p. 291.)

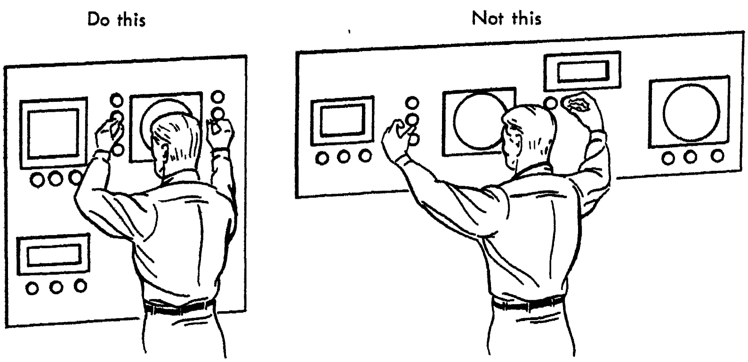


Fig. 16.19. Generalized illustration of desirable and undesirable arrangements of wall control panels. (Adapted from Morgan, *et al., op. cit.*, p 300.)

[27] For further discussion of this topic, see McCormick, *op. cit.*, and Morgan *et al., op. cit.*

HUMAN FACTORS IN JOB DESIGN

It has been suggested by Davis [28] that approaches to job design may be classified as: (1) process-centered (or equipment-centered); (2) worker-centered; and (3) the combination of these two. In the *process-centered approach*, jobs are designed by specializing activities or functions or by applying rational methods to determine minimum production time for different production methods. The *worker-centered* approach is represented primarily by job enlargement (which will be discussed shortly). Another worker-centered method is one in which a team of workers participates in deciding how the work will be divided among them. In the worker-centered approaches to job design, motivation is a central consideration. In the process-centered approaches, motivation is not.

Prevailing Considerations in Job Design Historically, and as of the present, jobs are overwhelmingly designed in terms of process-centered considerations. Davis, Canter, and Hoffman [29] report a survey that was carried out to determine the precepts or principles used by the responding companies in designing jobs. This survey indicated very clearly that those in industry who create, design, or establish jobs are guided almost entirely by the following: economic considerations or hypotheses, process considerations, time or space considerations, skills available and numbers of people available, tools and equipment required, union-management agreements, custom or tradition, and precepts such as specialization, repetitiveness, etc. These considerations are essentially predicated upon a mechanistic philosophy, with the primary criterion of immediate costs of production.

Motivational Considerations in Job Design Chapter 12 included a discussion of the research of Herzberg *et al.*[30] relating to variables associated with job satisfaction. It will be recalled that the variables that were associated with positive job satisfaction generally were variables that were related to the *job activities* as such (achievement, recognition, the work itself, responsibility, and advancement) rather than with the work situation. The worker-centered approach to job design that is characterized by Davis [31] is relevant in this context, since it is predicated on

---

[28] L. E. Davis, "The Concept of Job Design and Its Status in Industrial Engineering," in *Symposium on Human Factors in Job Design* (This symposium was held at the meetings of the American Psychological Association in New York, 1961; the papers are published by the Systems Development Corporation, Santa Monica, Calif., as Report SP–611, November 20, 1961).

[29] L. E. Davis, R. R. Canter, and J. Hoffman, "Current Job Design Criteria," *Journal of Industrial Engineering,* 6 (No. 2, 1955), 5

[30] F. Herzberg, B. Mausner, and B. B. Snyderman, *The Motivation to Work* (New York: John Wiley & Sons, Inc., 1959).

[31] *Davis, op. cit.*

the principle of designing jobs with consideration of the workers' motivation in mind. The objective of this approach would be to design jobs so that there is reasonable opportunity for employees to achieve positive job satisfaction from their work. But a worker-centered approach to job design cannot be justified exclusively on the grounds of satisfying the "human" requirements of jobs. It must also satisfy the technical organization requirements in an efficient, economical manner (which, incidentally, usually is the case).

Job Enlargement Job enlargement is the obverse of job specialization. As a worker-centered principle of job design, it is directed toward combining a variety of tasks or activities into a job. Job enlargement has been defined by Kilbridge [32] as ". . . the expansion of job content to include a wider variety of tasks and to increase the worker's freedom of pace, responsibility for checking quality, and discretion for method." The underlying assumption is that enlarged jobs will be intrinsically more satisfying to employees than highly specialized jobs. It is probable that there are many types of specialized work activities that, as such, cannot be intrinsically interesting to people. The proponents of job enlargement, however, would suggest that a meaningful *combination* of such activities could stimulate greater job satisfaction on the part of employees.

In discussing job design, Jasinski [33] refers to three basic types of factors: namely, organizational, technological, and individual. With respect to the technological and individual factors, in recent years a few social scientists have been concerned with man's social-psychological relationship to the technological processes with which he is associated. The current interest in job enlargement is a reflection of this concern.

Historical background. The current interest in job enlargement was in part initiated by Walker.[34] While there has not been a major shift in job design toward job enlargement during the subsequent years, there have been at least a handful of situations in which job enlargement has been carried out.[35]

Results from job enlargement. The reported effects of job enlargement programs include improved attitudes on the part of personnel,

---

[32] M. D. Kilbridge, "Reduced Costs Through Job Enlargement: A Case Study," *The Journal of Business of the University of Chicago,* 33 (No. 4, October, 1960).

[33] F. J. Jasinski, "Organization Change and the Job Design Process," in *Symposium on Human Factors, op. cit.*

[34] C. R. Walker, "The Problem of the Repetitive Job," *Harvard Business Review,* 28 (No. 3, 1950), 54–58.

[35] A few of these are listed below:

R. H. Guest, "Job Enlargement—A Revolution in Job Design," *Personnel Administration* (1957), 9–17.

L. E. Davis and R. Werling, "Job Design Factors," *Occupational Psychology* (London), 34 (No. 2, 1964), 109.

J. F. Biggane and P. A. Stewart, *Job Enlargement: A Case Study,* State University of Iowa, Bureau of Labor and Management, Research Series No. 25, July, 1963.

reduced turnover and absenteeism, increased productivity, and improved quality. While some job enlargement programs have been accompanied by moderate increases in production, job enlargement should not be viewed as a panacea for the production problems of an organization. It probably is meaningful, however, to call attention to the fact that production does *not* tend to decrease (as might be hypothesized on the basis of a "job-specialization" frame of reference).

One job enlargement case study will be summarized briefly. The case in question, reported by Biggane and Stewart,[36] dealt with the operations of assembling the water pump for an automatic washing machine at the Maytag Company. The pump includes 26 parts, and originally was assembled by five men on an assembly-line basis. In the enlarged method, one operator performed all assembly operations. This involved rearrangement of the work areas, and changes in the materials handling to move materials to and from the work areas. Some of the changes that followed the job enlargement were:

> Improved quality (reduction of rejects from 5.0 to 0.5 per cent).
>
> Cost reduction (direct labor costs reduced 16 per cent; materials handling and repair reduced 50 per cent); cost of equipment and installation recovered in less than six months.
>
> New operators do not affect production rate and earnings of others.
>
> Reduction in turnover.
>
> Improved housekeeping.
>
> Reduction in time for handling grievances and other matters usually associated with changes.

On the basis of this and two other case studies, Biggane and Stewart[37] conclude that "job enlargement offers definite opportunities to enhance the meaning of work through greater involvement of the operator, to favorably affect quality and cost, and to provide an opportunity for greater job satisfaction for the man on the job, and his supervisor."

Guidelines in job enlargement. At the present time, there are no particularly solid procedures, principles, or other guidelines to follow in job enlargement programs. While it will probably be a long time before research and experience provide the basis for any comprehensive principles and guidelines, Davis[38] has made some modest beginnings in this direction. In the meantime, the few case studies that have been reported probably can at least provide some ideas for the kinds of enlargement that might be feasible in other situations.

Discussion The job enlargement concept in job design seems essentially incompatible with the traditional method's analysis approach of the

---

[36] Biggane and Stewart, *op. cit.*

[37] *Ibid.*

[38] Davis, "The Concept of Job Design," *op. cit.*

industrial engineers that has tended to be focused on a mechanistic, work-specialization approach to job design. To some degree human factors engineering also is guilty of concentrating on the mechanics of human activities from a "micro" point of view—of simplifying work for people, and reducing it to fairly constrained, rather definitely programmed boundaries. The irrevocable shift in technology toward automation has contributed further to the process-centered bases of job design. On the other side of the coin, the experience to date with job enlargement does not justify an industry-wide, all-out effort to enlarge jobs in the interests of either industry as such or the personnel involved.

If one is interested in some rational approach to job design, where, then, does this leave us? Accepting some risks, we will offer some admittedly subjective observations. In the first place, while job enlargement has as yet not been supported by wide-scale experience, the evidence so far suggests two conclusions: it distinctly contributes to employee motivation; and it may be as productive as, or more productive than, job specialization in many job situations. In the second place, it is doubtful if a job enlargement approach ever could be justified on practical grounds as "the" basis for job design; as suggested by Nadler,[39] it cannot and should not become a whole program. In the third place, it is probable that the very nature of some work processes precludes the practical possibility of job enlargement, thus imposing some limits on the possible range of *types* of work activities that would be susceptible to such an approach

And, in the fourth place, it is suggested that the process-centered and worker-centered approaches to job design are perhaps not as incompatible as they might initially appear to be. In this connection, Davis[40] points out that operations planning takes place at two levels: namely, *task design* (to accomplish elements of operations), and *task combination* in which tasks are combined into jobs. Davis expresses the opinion that while industrial engineers cannot generally be criticized for the depth and intensity of their efforts in task design, they have not done well in task combination. It is primarily in the process of combining tasks into jobs that considerations of human motivation, job satisfaction, group behavior, and such variables come into play. Thus, at least in some job design situations, one can take advantage of a process-centered approach to task *design,* and a worker-centered approach to task *combination* (combining tasks into jobs).

A final comment should be added. In job design, it is manifest that the technical-organizational requirements must be fulfilled. But there is a growing feeling that perhaps greater emphasis should be placed on the

---

[39] G. Nadler, *Work Design* (Homewood, Ill.: Richard D. Irwin, Inc., 1963), p. 33.
[40] Davis, "The Concept of Job Design," *op. cit.*

human requirements of those engaged in the productive operations of our economy, especially in providing opportunity for achieving some reasonable personal satisfaction from work. There is some support for the notion that these two objectives are not intrinsically in opposition to each other, but that both might be achieved. As pointed out by Cantor,[41] however, there has been little attention devoted to the question as to what constitutes an effectively designed job, or to the development of principles and methods for designing jobs in terms of some accepted objective. These are matters of unfinished business to which both industrial psychologists and industrial engineers need to address themselves.

---

[41] R. R Cantor, "Human Factors in Job Design," in *Symposium on Human Factors in Job Design, op. cit.*

V

Human Errors and Accidents

The tendency toward error is a pervading human trait. Errors of various kinds can of course affect the quality of work people do, and also can contribute to injuries and fatalities. The sources of errors and accidents, however, are not entirely those of the individuals in question; the nature of the work activities, the design of the equipment, the procedures, and the work environment—these and other variables can have some bearing upon the frequency and nature of errors and accidents. This section deals with some of the aspects of errors and accidents in industry.

17

Human Error

Human errors in work activities are so pervading that they have given rise to such literary expressions as "to err is human, to forgive divine." In many phases of work, errors are a nuisance, impose administrative problems (such as inspection), and are costly, but these burdens may not be critical. In some types of situations, however, errors may result in such consequences as to justify virtually any effort to eliminate them. Where errors can be extremely critical (such as in space travel), there is talk of "error-free" performance, and efforts are made virtually to eliminate the possibility of error. By and large, however, in terms of practical considerations, it is probably not reasonable to think in terms of eliminating human error. It is feasible, however, to consider the possibility of reducing error to acceptable levels.

Human errors, of course, can be manifested in many specific ways, but in general their consequences are reflected in poor quality of products, increased costs of production, damaged materials and equipment, and personal injuries and death.

NATURE OF HUMAN ERROR

To consider any action toward the reduction of error to reasonable levels, it is necessary to have relevant information about the specific types of errors in question, and about the human behaviors and situational variables that are associated with them. In fact, in considering the types of errors, there is the basic question of what really constitutes an error. An operational definition of human error has been proposed by Peters[1] as follows: "Any deviation from a previously established, required, or expected standard of human performance that results in an unwanted or undesirable time delay, difficulty, problem, incident, malfunction, or failure." As pointed out by Rook,[2] human error is somewhat synonymous with poor workmanship. Since workmanship frequently can be considered as varying along some continuum, it may be necessary to characterize that degree of workmanship that is considered unacceptable in terms of some appropriate standard. In some circumstances, however, the consequences of an error are so clear cut that there is no question of degree. Assuming that, in one way or another, it is possible to characterize errors in a work situation, it is useful to determine, if possible, the human behaviors and the situational variables associated with the type of error in question. While it is usually difficult to establish basic cause-and-effect relationships, such information frequently can serve as the basis for corrective action.

Categories of Error One system of classifying errors has been proposed by Rook.[3] The system was developed on the basis of an analysis of 23,000 production defects that occurred in industrial production operations. This system provides for classifying errors in terms of two dimensions, with three categories in each (making a total of nine categories).

Conscious awareness or intent. One of these dimensions is related to the conscious awareness or intent of the individual, as follows:

A. *Intentionally performed.* (Errors that result from acts performed at the conscious level with the belief that the result of the act will be within the correct limits of performance, but the results of which are outside acceptable limits.)
B. *Unintentionally performed.* (Errors that result from acts which are not intentionally performed; the act "just happened.")
C. *Omitted.* (Errors that result because of failure to do something.)

---

1 G. A. Peters, "Human Error and 'Goof Proofing,'" paper presented at Product Assurance Symposium, American Society for Quality Control, San Fernando Valley Section, Glendale, Calif., October 20, 1962.

2 L. W. Rook, Jr., *Reduction of Human Error in Industrial Production* (Albuquerque, New Mexico: Sandia Corporation, Technical Memorandum SCTM 93-62 [14], June, 1962).

3 *Ibid.*

Behavior components. The second dimension of the classification system was predicated in part on a classification developed by Payne and Altman[4] in which typical "behavior components" are placed in one of three classes, namely, *input* behaviors, *mediating* behaviors, and *output* behaviors. Rook postulated that errors could generally be attributed to one (or in some cases more) of these behavior components.

Cross-classification of dimensions. The cross-classification of these two dimensions is given in Table 17.1. Granting some difficulty in precise classification, it can be seen that a behavior that results in an error can generally be characterized in terms of whether it was due to *input* (I), *mediation* (M), or *output* (O). Within each of these it could also be characterized as *intentional* (A), *unintentional* (B), or *omitted* (C).

Table 17.1

SYSTEM OF ERROR CATEGORIES *

Conscious level or "intent"		*Behavior component*	
	Input (I)	*Mediation* (M)	*Output* (O)
The act is			
A. Intentionally performed	AI	AM	AO
B. Unintentionally performed	BI	BM	BO
C. Omitted	CI	CM	CO

* *From Rook, June, 1962,* op. cit., *Table IV.*

Examples of system. An example or two of the classification system would be in order. A defect which consists of the reversal of two wires in an assembly, one blue and the other purple, probably could be attributed to a confusion between the two colors (an error in information input), but the worker intended to do what he did (even though he was confused on the input). This error would then be classified as an intentional act (A) by the one system, and as due to input (I) by the other system. In case a solderer burns the insulation of the wire, it could be suggested that the motor behavior which brought the soldering iron into contact with the insulation was unintentional (B), of an output (O) type. Obviously, some categories would have few, if any, errors, but this does not affect the possible utility of the basic system.

Basic Sources of Errors While it usually is difficult to isolate the "real" causes of individual errors, it is reasonable to postulate that (theoretically) they can be attributed to two primary classes or sources or a combination of them: *individual* or personal factors, and *situational* factors. A third factor, *training*, might be added to this list, although our subsequent discussion will deal largely with the first two.

[4] D. Payne and J. W. Altman, *An Index of Electronic Equipment Operability*, American Institute for Research, Report AIR-C-43-1/62, 1962.

Where there are several, or many, individuals doing the same work in the same situation, there may be consistent differences in the error-rates of different individuals. One may not know, however, what specific *types* of individual differences give rise to these systematic differences in error-rate. On the other hand, the situation may have an important influence on the error-rate—high, or low, as the case may be. Situational variables can range over a wide gamut, such as the design of the equipment or tools, the work layout, methods of work, duration of work periods, and physical environment.

ANALYSIS AND USE OF ERROR DATA

From the above discussion, it follows logically that possible corrective action should depend upon knowledge about the major sources of human error. If there is evidence that errors vary widely among *individuals* who are engaged in similar work, the possible appropriate actions lie in the direction of personnel selection, training, and motivation. In connection with personnel selection, for example, it would be appropriate to identify, through regular validation procedures, those predictors (tests, personal data, etc.) that can be used to predict which individuals would generally have low error rates. (Since such validation procedures as well as training and motivational considerations were discussed in earlier chapters, they need not be repeated here.)

If, on the other hand, there is evidence that error rates are associated with *situational* variables, the problem is that of identifying the particular feature(s) of the situation that are related to high error rate in order to do something about them. An overly simplified and generalized representation of the relationship between *individual* and *situational* variables as related to error rate is shown in Fig. 17.1. Some actual examples of both types of variables as related to errors will be given later, but at the moment it would be pertinent to point out that actual error data—of whatever type—preferably should be quantified, such as in terms of probabilities (as illustrated in Fig. 17.1).[5]

Methods of Identifying Sources of Human Error It has been suggested by Peters[6] that there are four principal methods that can be used to identify sources of human error. These methods, in somewhat modified form, are described below.

[5] One method for quantitatively evaluating the contribution of human error to the degradation of product quality or performance has been developed by Rook, June, 1962, *op. cit.* The quantification of error (human and/or machine) in terms of probabilities is related to the concept of the *reliability* of man-machine systems as expressed in terms of probabilities of successful operation or performance (in this sense, *reliability* has a different meaning than as used elsewhere in this text).

[6] Peters, *op cit.*

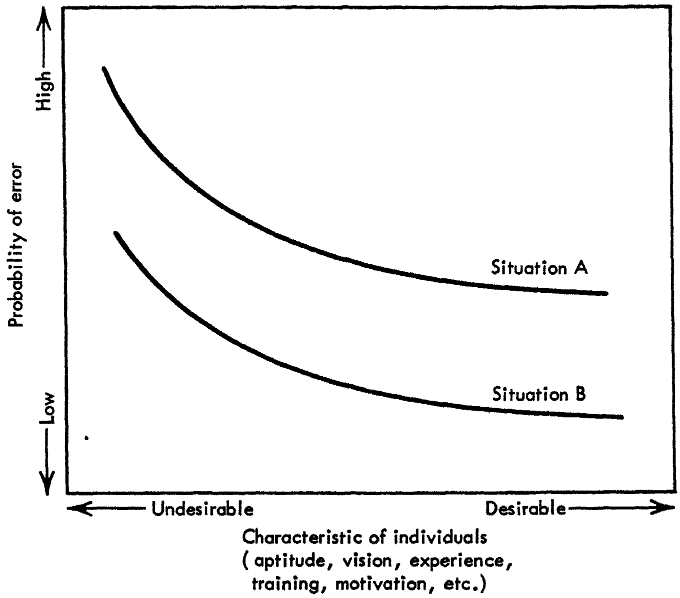


Fig. 17.1. Generalized relationship of situational variables and of individual characteristics as related to error rate (probability of error). Situations A and B might be differences in equipment design, methods, work periods, environments, etc. (Adapted from Rook, June 1962, *op. cit*, Fig. 1.)

Data collection and analysis. In the case of an on-going operation, error data can be obtained through reported malfunctions, system failures, inspection reports, equipment logs, accident records, personal injuries, etc. Such data can be summarized in various ways, such as by the classification scheme shown in Table 17.1. Such a summarization (especially if accompanied by appropriate statistical analyses) can start to give cues as to possible sources of error. For example, wide variation in error rates among individuals on the same job (and in the same situation) can suggest further analysis of individual variables associated with error rates. In turn, significant differences in error rates between or among different situations (such as different environments, work methods, etc.) focus more on situational factors.

Direct observation. Observations of work processes on a continuous or sampling basis sometimes can aid in identifying situational variables that are conducive to errors.

Systems analysis. Where a new, or modified, system is being planned, an analysis of the job operations and operations sequence may serve to locate potential problems. Such analyses sometimes are made on the

basis of design drawings or blueprints, with the view toward identifying potential sources of error before the system is developed.

Simulation. The experimental use of prototypes, mockups, or other forms of simulation may help to identify potential sources of error in the design and development of systems.

The last three methods would be primarily useful in identifying situational variables that may be associated with errors. By the first method (data collection and analysis) it may be possible to determine the relative influences of individual and of situational variables in error rates.

Discussion The variations in human error are so integrally related to the nature of the human activity in which they arise that it is not feasible to discuss or illustrate very many contexts in which they occur. For illustrative purposes, human error will, therefore, be discussed in the following contexts: monitoring tasks; inspection operations; and accidents.

MONITORING TASKS

The trend toward greater and greater mechanization and automation is increasing the number of jobs in which a major function is that of monitoring an operation or process. In typical monitoring tasks (sometimes called vigilance tasks) the monitor's function is that of giving his attention to the operation in order to identify circumstances or events that require some action (response) on his part. In general, such tasks are characterized by prolonged periods of time accompanied by infrequent stimulus events that are to be identified.

A primary requirement for a monitor is that of correctly identifying all, or most, of the events that should require his action. The input relating to these events may be presented to the monitor by various displays (such as cathode-ray tubes, dials and gauges, instruments of various kinds, auditory signals, etc.), or they may be observed or detected directly (such as noticing a change in the sound of a machine). Basically, then, this requires the correct identification of all relevant stimuli (there should be no misidentifications of wrong stimuli, nor omissions of identification of relevant stimuli). In studies of monitoring performance various criteria of performance have been used. These include: (1) errors of omission (failure to respond to a signal); (2) errors of commission (responding to a "wrong" stimulus); and (3) response lag. Of these, errors of omission are most common, and undoubtedly represent the most typical error experienced in this type of task. Most monitoring studies involve the presentation of occasional signals (usually visual or auditory), with instructions to the monitor to make a specific response when he detects the signal.

While there are, of course, individual differences in the ability to monitor tasks, the primary concern in monitoring activities has been in connection with the effects of situational variables on monitoring performance. A couple of such variables will be discussed.

Duration of Monitoring Periods The duration of the monitoring period has been found to have an important bearing on monitoring performance. In a number of experiments it has been found that the percentage of signals detected typically goes down with time. An example of this comes from a study by Bergum and Lehr [7] in which 20 male subjects sat in a booth watching a circular panel with 20 one-half-inch red lights which illuminated in sequence at a rate of 12 rpm. The "signal" to be detected consisted of the failure of a light to illuminate in its normal sequence (the signal rate was 24 per hour). Some subjects monitored this panel for 90 minutes continuously. Their performance (expressed as percentage of signals detected) is shown in Fig. 17.2 as line

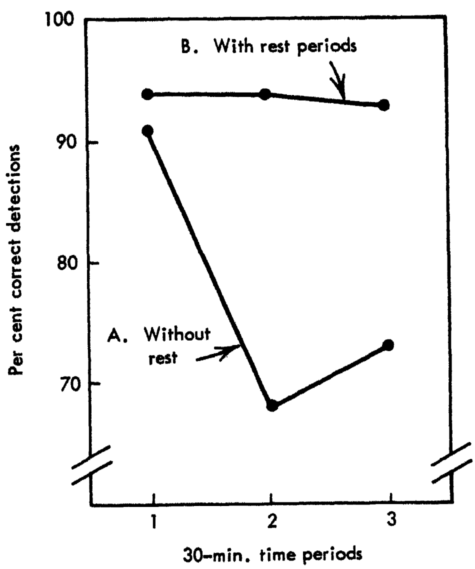


Fig. 17.2. Percentage of correct detections of visual signals for 90 minutes of continuous monitoring, and for 90 minutes with two 10-minute rest periods. (Adapted from Bergum and Lehr, *op. cit.*, Table 1.)

A. That figure also shows, for comparative purposes, the performance of another group who were given 10-minute rest periods at the end of the

[7] B. O. Bergum and D. J. Lehr, "Vigilance Performance as a Function of Interpolated Rest," *Journal of Applied Psychology*, 46 (1962), 425–427.

first 30 minutes and after the next 30 minutes of monitoring. In this, and other studies, it has been found that occasional rest periods (such as every 20 or 30 minutes) are helpful in maintaining a reasonable level of performance.

Frequency Rate of Signals Another factor that has frequently been found to affect monitoring performance is the frequency of the signals, with performance *decreasing* with *low* (as opposed to high) signal rates. This is illustrated, for example, by the results of a study by Deese [8] in which the signals per hour for various experimental periods were 10, 20, 30, and 40. The results of this study are shown in Fig. 17.3, and indicate quite clearly the improved performance with higher (rather than lower) signal rates. Relatively comparable results are reported from numerous other studies, such as the one by Kappauf and Powe,[9] who

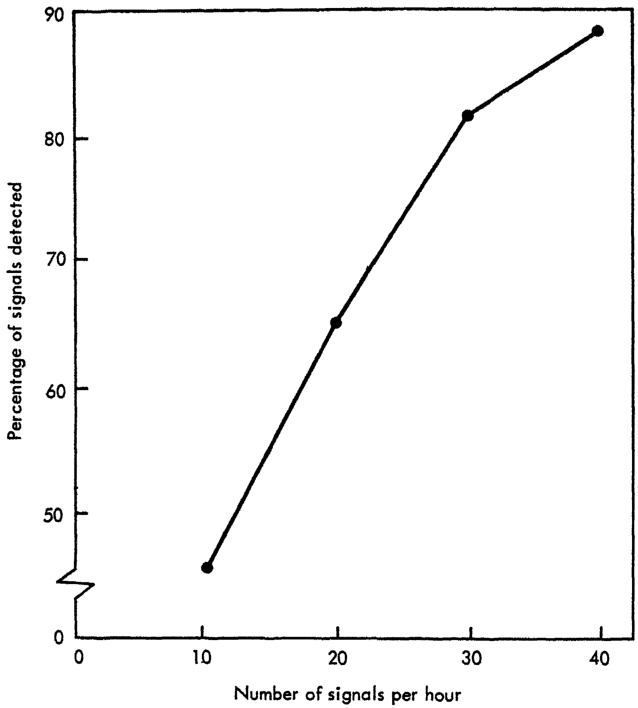


Fig. 17.3. Percentage of signals detected for varying signal rates. (Adapted from Deese, *op. cit.*)

---

[8] J. Deese, *Changes in Visual Performance After Visual Work*, USAF, Wright Air Development Center, Technical Report 57-285, 1957.

[9] W. E. Kappauf and W E. Powe, "Performance Decrement on an Audio-visual Checking Task," *Journal of Experimental Psychology*, 57 (1959), 49–56.

experimented with signal rates of 8, 20, 40, and 80 per hour in an audio-visual digit checking task. It should be noted that while lower detection rates for low signal rates have been found in many studies, this is not universally the case. In the study by Bergum and Lehr,[10] for example, there was not much performance difference between rates of 24 and 6 signals per hour. They point out that such performance differences typically have been found where the differences in signal rates are very marked (of the order of 4 to 60), but not where the differences in signal rates are low (such as by a factor of four or less).

Where the signal rate is so low as to produce a performance decrement, the introduction of artificial signals has sometimes been found to be useful in maintaining a satisfactory level of signal detection. While this procedure has not invariably been found to be useful,[11] it is a procedure that is at least worthy of consideration as a means of maintaining monitoring performance.

Discussion Aside from the variables that have been discussed above, other situational variables have also been found to be related to the detection of signals in monitoring tasks. These include intensity of signals, the use of two sensory modalities in combination, knowledge of results, and supervision.

Various theories of vigilance have been postulated, these generally being related to such concepts as conditioning and reinforcement, expectancy, sensory variation, arousal, and motivation. While certain of these seem more adequately to account for monitoring behavior than do others, there is as yet no consensus regarding any single theory as being completely adequate.[12]

Certain practical approaches to the achievement of reasonably adequate levels of performance in monitoring tasks have been given or implied above. Considering the results of available research, a number of principles have been suggested by Bergum and Klein [13] for the more effective design of man-monitored systems. These are given below (with minor adaptations):

---

10 Bergum and Lehr, *op cit.*

11 B. O. Bergum and D. J. Lehr, *Vigilance Performance as a Function of Task and Environmental Variables,* Human Resources Research Office, Research Report 11, May, 1963.

12 For discussions of various theoretical positions (and other aspects of vigilance) the reader is referred to the following sources:

B. O. Bergum and I. C. Klein, *A Survey and Analysis of Vigilance Research,* Human Resources Research Office, Research Report 8, November, 1961.

D. N. Buckner and J. J McGrath, Editors, *Vigilance; A Symposium* (New York: McGraw-Hill Book Company, 1963).

J. P. Frankmann and J. A. Adams, *Theories of Vigilance,* USAF, Air Force Command and Control Development Division, Air Research and Development Command, Technical Note AFCCDD-TN-60-25, April 30, 1960.

13 Bergum and Klein, *op. cit.*

1. Visual signals should be as large in magnitude as is reasonably possible. This includes size, intensity, and duration.
2. Visual signals should persist until they are seen (or otherwise detected), or as long as is reasonably possible.
3. In the case of visual signals, the area in which a signal can appear should be as restricted as possible.
4. Although "real" signal frequency often cannot be controlled, where possible it is desirable to maintain signal frequency at a minimum of 20 signals per hour. If necessary, this should be accomplished by introducing artificial (noncritical) signals to which the operator must respond.
5. Where possible, the operator should be provided with anticipatory information. For example, a buzzer might indicate the subsequent appearance of a critical signal.
6. Whenever possible and however possible, the monitor should be given knowledge of results.
7. Noise, temperature, humidity, illumination and other environmental factors should be maintained at optimal levels.
8. The system should be so designed that operators do not work in isolation from other individuals.
9. Whenever possible, individual watches should not exceed 30 minutes.

INSPECTION OPERATIONS

Various types of inspection operations represent another class of work activities in which human errors should be minimized as much as possible. (While inspectors typically are inspecting the work of others, they are not immune to making errors themselves!) It should be noted, however, that some inspection operations are somewhat comparable to monitoring activities, since they may involve prolonged periods of time during which attention is concentrated, and infrequent stimuli to which responses are to be made. Where inspection operations are similar to monitoring tasks, much that has been said about monitoring tasks also would apply.

Obtaining Data on Inspection Errors Since a rational approach to the reduction of errors in inspection requires the availability of relevant information, some examples will be given of inspection operations in which relevant data have been collected and analyzed. Because of the problems associated with data collection from on-going inspection activities, these examples generally illustrate the use of job samples. Job samples are actually examples of the method of simulation mentioned earlier.

Tin-plate inspection. One of these examples concerns the job of tin-plate inspection, called assorting.[14] Since this job will be referred to

[14] J. Tiffin and H. B. Rogers, "The Selection and Training of Inspectors," *Personnel*, 18 (1941), 14–31.

later, it will be described briefly. The operation is essentially an inspection for appearance that is made while the inspector turns the sheets of tin-plate from one stack to another. As the sheet is turned, the inspector makes a decision from the appearance and feel of the sheet whether it is a prime or a second or contains one of a number of possible defects. In the plant studied, the work is done by women assorters. Supervisors have generally felt that inspectors on this job do not reach their maximum performance until they have had approximately six months of experience. The inspectors are paid on a straight hourly rate.

The mechanics of wrapping and shipping made it impractical to reinspect samples of tin-plate in order to evaluate the performance of the inspectors. Under the circumstances, a job sample test was used. In its final form it consisted of a coded stack of 150 sheets. This coded stack was made up of 61 prime sheets (sheets satisfactory in every respect) and second sheets (sheets containing a minor surface blemish or uneven coating of tin), 30 sheets containing appearance defect No. 1, 26 sheets containing appearance defect No. 2, 13 sheets containing appearance defect No. 3, and 20 sheets containing a weight defect. The prime and second sheets were included in a single category for the purpose of the experiment because with repeated assortings the prime sheets tend to become scratched and thus to become seconds.

The defects will be referred to by number rather than by name because it is the method of measuring the accuracy of the inspectors rather than their accuracy on specific defects that is of primary importance. The 150 sheets were numbered in random sequence. As the inspectors assorted the sheets, they called aloud their judgment of each sheet, whether it was a prime or a second or contained one of the defects, and, if so, which one. The job sample test was given without a time limit in order to make it possible for each girl to demonstrate her highest level of inspection ability. Each of the 150 inspectors who participated, however, was timed with a stop watch.

The coded stack was scored for each girl according to the total accuracy on the 150 sheets and also for the accuracy on the specific defects. This procedure resulted in five specific measurements of accuracy for each girl. It also resulted in a measurement of the reliability of each of the five methods of scoring the coded stack test. The five resulting measurements with their respective reliabilities are summarized in Table 17.2.

The reliability figures given in Table 17.2 were obtained by correlating correct inspection on odd *vs.* even items for each defect. The reliability indicates the extent to which repeated or duplicate measurements of each girl by means of the coded stack test would result in the same score for her for the defects in question. It may be seen from Table 17.2 that the reliabilities vary from .68 to .90.

The results of the coded stack test as administered to the 150 girls

Table 17.2

THE FIVE MEASUREMENTS OF ACCURACY YIELDED BY THE CODED STACK, WITH THE RELIABILITY OF EACH

Method of scoring	*Reliability*
Mixed sheets	+.90
Appearance defect No. 1	+.86
Appearance defect No. 2	+.87
Appearance defect No. 3	+.68
Weight defect	+.74

are summarized in graphic form in Fig. 17.4. The six curves shown in this figure are frequency distributions obtained from the 150 inspectors.

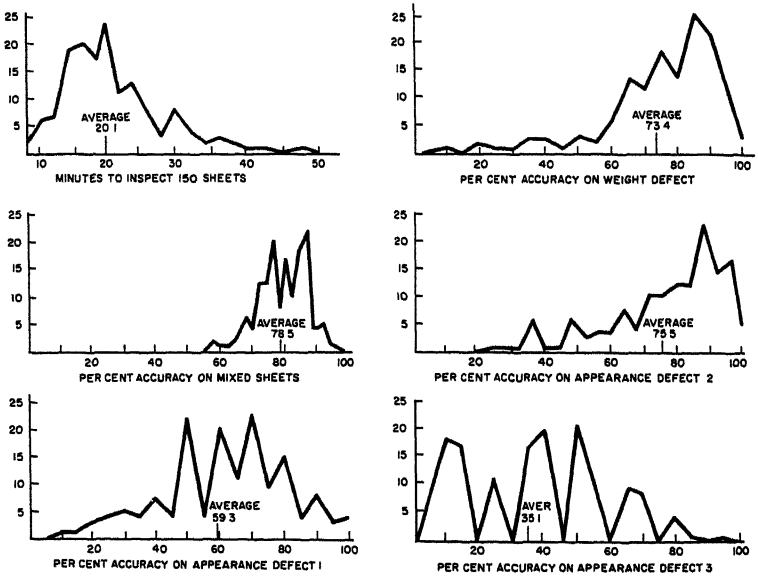


Fig. 17.4. Distributions of results on a coded stack test of 150 tin-plate inspectors.

The time curve in the upper left-hand corner indicates the number of girls completing the 150 sheets of tin-plate in each of the times specified along the base line. The times ranged from 8 minutes to 48 minutes, with an average of 20.1 minutes. It will be seen that the average

accuracy for the mixed sheets was 78.5, with the averages for the specific types of defects ranging from 35.1 per cent to 75.5 per cent. The wide variation among individuals is reflected by the spread of the distributions.

The next question investigated was the relation between speed and accuracy in detecting the different types of defects. The results for this part of the investigation are summarized in graphic form by the curves in Fig. 17.5.

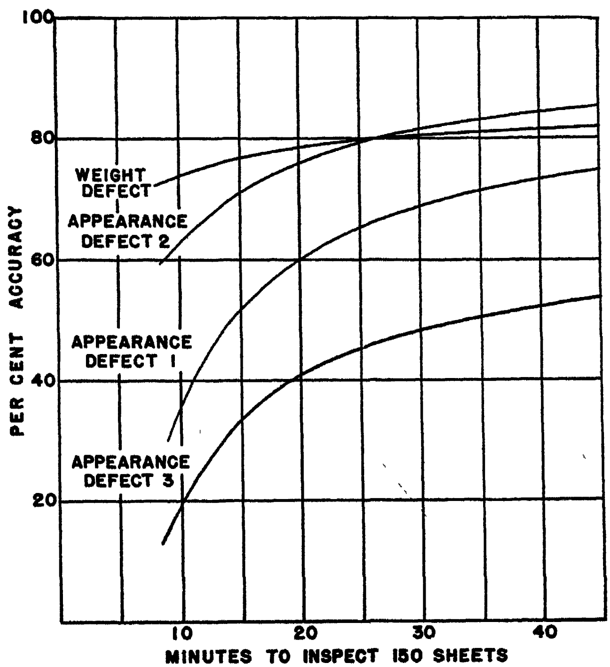


Fig. 17.5. Relationships between speed of inspection and accuracy in detecting four defects.

It is interesting to note that in detecting off-weight sheets the accuracy is about the same whether the test is completed in 10 minutes or in 40 minutes. In other words, accuracy in detecting off-weight sheets for the 150 girls tested is not appreciably affected by speed of inspecting within the time limits found for these girls. To some extent the same general situation is true for appearance defect 2, although in detecting this defect we note a beginning tendency for accuracy to decrease with increased speed. This tendency is still more pronounced in the case of appearance

defects 1 and 3. For example, we find that when the stack is assorted in 10 minutes, only 20 per cent of the sheets containing appearance defect No. 3 are spotted and only 35 per cent of those containing appearance defect No. 1.

Information such as that obtained by this job sample test would be useful in various ways in efforts to improve the accuracy of the inspection operation. (Further mention of this study will be made later.)

Accuracy in bottle inspection. In the manufacture of bottles, inspectors (called selectors in this industry) examine the bottles before they are packed for shipment. One glass bottle company became concerned about the accuracy of the selectors and asked an industrial psychologist to work on the problem.[15] As a first step, a job sample consisting of 288 bottles was assembled. These 288 bottles were carefully chosen by a quality control management committee so that they contained a few of all common defective bottles. The composition of the 288-bottle job sample consisted of the following:

No. of bottles	*Type of bottle (by type of defect)*
34	marked
48	checked shoulder
30	warped finish
20	checked ring
26	neck tear
130	packable bottles (no defect)

The labels in the above column describe well known and common defects in the bottle industry. Since it is difficult to describe them without actual examples, the investigation will be described without a further description of the defective bottles. Forty of the experienced selectors were each asked to sort the 288-bottle job sample and report how each bottle would be classified. The results obtained from each selector were then scored for each type of bottle and reliabilities of the job sample sub-scores were computed. The resulting reliabilites are summarized below:

Type of bottle	*Coefficient of reliability*
all bottles	.85
all defective	.85
all packable	.85
marked	.81
checked shoulder	.47
warped finish	85
checked ring	.71
neck tear	.67

15 H. C. Coe, *Studies in Glass Bottle Selection,* Ph.D. Thesis, Purdue University, 1956.

The reliabilities were all reasonably satisfactory except for the checked ring, neck tear, and (particularly) the checked shoulder defects. The results suggest that even these experienced selectors were not sure of the nature of some of these particular defects.

The fact that the best selectors were more stable in their judgments, however, is shown by the following summary which compares the accuracy of the four best selectors with a random group of 40 (all experienced) on each of the bottle types.

	Per cent accuracy	
Type of bottle	*Four best selectors*	*40 other selectors*
all bottles	93%	85%
all defective	93%	82%
all good	93%	90%
marked	93%	75%
checked shoulder	97%	95%
warped finish	91%	67%
checked ring	98%	92%
neck tear	87%	75%

The fact that even the 40 random selectors attained 95 per cent accuracy on the checked shoulder bottles suggests that this defect is quite easy to identify.

The use of this 288-bottle job sample enabled the company to identify the best selectors as well as the particular types of defect that were most frequently missed by many selectors. The latter information was used by the training department in setting up a training program for these employees. The selectors were brought together in small groups and shown several examples of these "hard-to-identify" defective bottles. Such training gave them the information needed to be more accurate on the job.

Use of precision instruments. Other examples of job sample tests with inspectors include a study by Lawshe and Tiffin.[16] In particular, this investigation dealt with the measurement of metal parts with precision instruments. Many precision instruments used in machine shops require that the operator judge the "feel," "tension," "drag," or other characteristics; yet, in spite of all that is known about the variability of human judgments, little attention has been given to the importance of such variability as it may affect the accuracy of measurements.

One of these examples is from a plant engaged in the manufacture of variable-pitch propellers for aircraft. Approximately 200 people were employed in the inspection department. Each of the various classifications of inspection jobs was analyzed to determine what precision instruments

---

[16] C. H. Lawshe and J. Tiffin, "The Accuracy of Precision Instrument Measurement in Industrial Inspection," *Journal of Applied Psychology,* 29 (1945), 413–419.

were used on the job, and what tolerances were demanded. Frequency counts were then made to determine which instruments or combination of instruments were used in the largest number of classifications and by the largest number of employees. On the basis of this count, 20 instruments and combinations were chosen as being most important in this particular plant.

A room was set aside as a dimensional control laboratory and 20 booths or inspection stations were set up. Each booth was numbered, and in each were placed one of the 20 instruments, a standard part from the plant, and a simplified working drawing that indicated one dimension to be measured with the instrument provided. When an employee entered the room, the attendant determined his job classification and provided him with an appropriate work-sheet for each of the stations containing work samples from his job. Each employee was tested on only those instruments that he used on his particular job. He was encouraged to make five measurements and then to record his best judgment concerning the dimension. The readings thus obtained were compared with so-called "true" dimensions, which were determined by means of ultra-precision instruments in combination with Johanssen blocks. Instruments utilized in the performance testing were checked and adjusted periodically to insure constancy.

Results obtained at 11 of the 20 stations are presented in Fig. 17.6. The particular stations selected for illustration were chosen in terms of general familiarity with the instruments used and not because of any peculiarity in the findings; they are truly representative. The accuracy in the use of the different measuring instruments ranges from a low of 9 per cent to a high of 66 per cent. Even the highest values would be below reasonably acceptable levels, thus reflecting a rather seriously high error rate.

A similar investigation was carried out with a group of 45 toolmakers in another plant. Each man measured each of 19 parts with a standard vernier micrometer, each part being measured independently 5 times; the measurement recorded was the best judgment he could then make of the "true" measurement on the basis of these trials. After all of the readings by all men had been completed, the parts were measured with ultra-precision instruments and Johanssen blocks in order to obtain the closest possible approximations to the "true" dimensions against which to compare the measurements made by the men. A frequency distribution of the per cent of toolmakers who measured different numbers of the 19 parts to within $\pm .0001$ inch of the "true" dimension is shown in Fig. 17.7. The percentages of toolmakers who were able to measure various parts within specified tolerances were similar to those cited above for the inspectors. The results further showed that the larger the part, the less accurate the measurement.

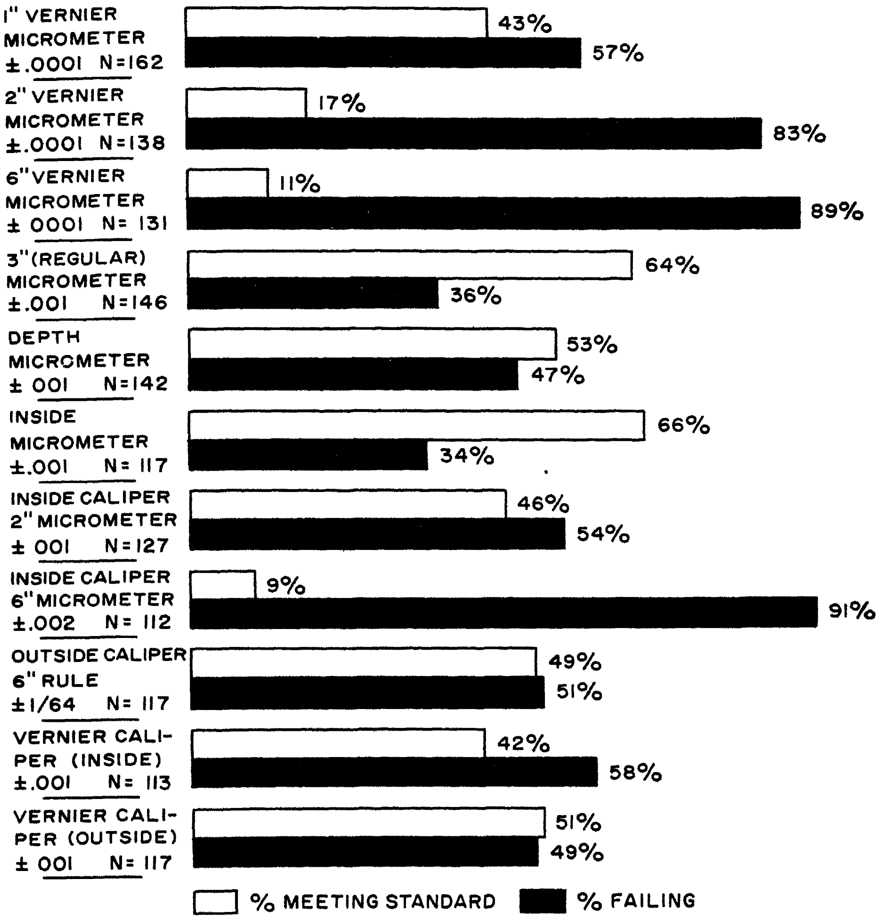


Fig. 17.6. The percentage of inspectors passing and failing various precision measuring instrument performance tests in an aircraft propeller plant. The open bars indicate the percentage meeting the standard and the solid bars indicate the percentage failing.

Another study of accuracy in using a micrometer was made by Evans.[17] Evans' results substantiate the findings of Lawshe and Tiffin[18] in that errors in micrometer measurements made by skilled workers are considerably larger than is assumed by most authorities in the field. He also found that the trade apprentices covered in his study were as ac-

[17] R. N. Evans, "Training Improves Micrometer Accuracy," *Personnel Psychology*, 4 (1951), 231–242.
[18] Lawshe and Tiffin, *op. cit.*

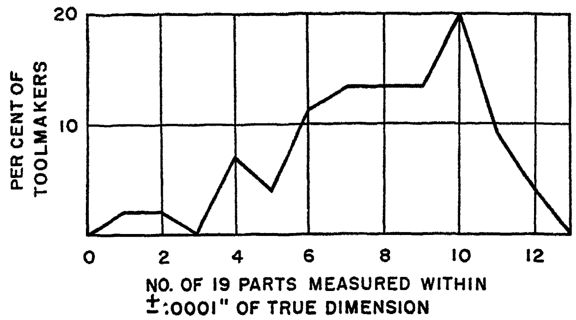


Fig. 17.7. Distribution of percentage of toolmakers who measured different numbers of 19 parts to within ± .0001 of an inch with a standard vernier micrometer.

curate as journeymen, and that there was no significant relationship in the case of either group between accuracy and age, amount of experience with the company, or length of time on the present job.

Although studies such as those reported above suggest that many errors are made in inspection operations, management need not throw up its hands in despair. As indicated earlier, sometimes actions can be taken to reduce human errors through personnel selection and training, and through the development of desirable work situations, such as proper methods.

Discussion In obtaining data on human accuracy in inspection operations, it is particularly desirable to do so in a reasonably objective manner (such as illustrated above) rather than to depend upon supervisors' ratings of the accuracy of inspectors. This probably can be attributed to the fact that supervisors frequently do not have a very objective basis for judging the inspectors' performance.

The authors can report an unpublished example of how ratings failed to identify the best inspectors from some work done with a well-known roller bearing company. At the time the work was done, roller bearings were subjected to an "appearance" inspection before being shipped to the customers. The company had been using a somewhat standard type merit rating plan to appraise the competence of the inspectors. The form required that the inspectors be rated by their supervisors on several characteristics. The sum of the ratings on these characteristics was considered a valid indication of job performance. One of the several specific characteristics on which ratings were made was "quality" of work.

A job sample of roller bearings was assembled and each inspector

was asked to inspect and report on the bearings in this sample. The results of the job sample inspection were then compared with ratings of the inspectors that had been made by their supervisors.

The results of this investigation are summarized in Tables 17.3 and 17.4. Table 17.3 shows that inspectors who were rated highest (ratings of 80–89) achieved only 75 per cent accuracy on the job sample, while those rated lowest (40–49) achieved a job sample accuracy of 79 per cent. There was no relation between ratings and job sample accuracy for the inspectors rated in the intervening rating categories. Table 17.4 shows that the average accuracy on the job sample of the inspectors rated "Good" was not appreciably better than of those rated "Poor," and the accuracy of those rated "Fair" is slightly, but not appreciably, better than any other of the rated groups.

These results, along with similar results from other studies on the relation between ratings and accuracy of inspectors, show that one should not use ratings as an indication of job performance for the work of inspectors. Factors other than the quality of work too often influence the ratings given to the employees. A job sample of the types described in the preceding discussions is a far more valid way to measure job performance in these instances.

Table 17.3

RELATION BETWEEN OVER-ALL MERIT RATING AND ACCURACY ON JOB SAMPLE OF ROLLER BEARING INSPECTORS

Total points on merit rating	*Average per cent correct on job sample*
80–89	75%
70–79	82%
60–69	78%
50–59	75%
40–49	79%

Table 17.4

RELATION BETWEEN RATINGS ON QUALITY AND ACCURACY ON JOB SAMPLE OF ROLLER BEARING INSPECTORS

Rating	*Average per cent correct on job sample*
Good	77%
Average	79%
Fair	84%
Poor	68%

There is one caution mentioned by Thomas and Seaborne[19] that should be kept in mind in this connection. They point out that much evidence has been produced to show that industrial inspectors are often inconsistent both with each other and with their own previous performance. They explain these inconsistencies by the fact that the perceptual "set" of the inspectors varies from time to time. In other words, if word gets around that a very important job is under way, many of the inspectors will do a very careful job, while if the job is more or less ordinary, they will tend to be a little less careful or perhaps even be careless. In the light of these observations by Thomas and Seaborne, one should emphasize the importance of the task when a job sample is being used. When this is carefully done, one will learn from the job sample how accurate the inspectors can be when they are working under an optimal "set."

Having illustrated a few examples of the use of job samples in obtaining data on human errors in inspection processes, let us now turn our attention to certain possible methods of reducing such errors.

Methods of Inspection In the development of methods of inspection, it is, of course, desirable to adopt whatever procedures take the best advantage of relevant human sensory and perceptual skills; this may, in some cases, require appropriate experimentation. A few examples of inspection methods that take advantage of such human skills will be given.

Methods of inspecting tin-plate. The tin-plate inspection study, previously mentioned, had various objectives. One of these was that of developing the best method of inspection. For this purpose motion pictures were taken, at 1,000 frames per minute, of twelve inspectors selected from those who had taken the coded stack test. In this group were two inspectors who had been rated fast and accurate, two rated fast but inaccurate, two rated slow and accurate, two rated slow and inaccurate, and four rated average in both speed and accuracy according to the test-stack data. The pictures were taken in the shop under normal working conditions, and because of the general noise and shop activity, the operators were not aware of the exact moment at which the pictures were taken.

Obviously the pictures did not reveal the accuracy with which the inspector detected the defects. They did show, however, the normal speed of the operator, the relative body activity or exertion, the various methods used in grasping and moving the sheets, and the disturbance in the rhythm whenever a defective sheet was found.

---

[19] L. F. Thomas and A. E. M. Seaborne, "The Social-Technical Context in Industrial Inspection," *Occupational Psychology,* 36 (1962), 36–43.

An analysis of these films revealed, in general, that most inspectors followed the movement of the sheets with their eyes, as illustrated by the left-hand pictures of Fig. 17.8. A few inspectors, however, followed a different eye-movement pattern; they inspected the sheets as they remained on the table rather than as they were being moved. These inspectors generally were the ones whose accuracy on the job sample was highest, thus suggesting that the method which they used might be a better one. This hypothesis has support from research evidence that the eyes can see a motionless object more accurately than one that is in motion. This method actually was adopted, and all of the inspectors were given appropriate training in the method.

The use of examples as standards. It has been rather clearly demonstrated in psychological investigations that people generally can make more accurate judgments when these judgments are *relative* rather than *absolute*. In other words, it is usually easier for people to judge amounts or degrees of qualities, traits, or characteristics of objects when *comparing* such objects with each other, than when attempting to judge such objects individually. This principle has applicability to inspection operations. The use of so-called "limit samples" illustrates the recognition of this principle. A limit sample is a sample of a product that is just barely acceptable in terms of inspection standards; it represents the "limit" of acceptability. Any product that is more defective than the limit sample therefore should be rejected. An inspector with a limit sample compares products with this sample, accepting them if they are equal to, or better than, the sample, and otherwise rejecting them. Such a comparison usually results in a more adequate judgment than when the inspector relies on a "memory image" of the degree of a defect that is acceptable *vs.* not acceptable.

An example of this approach is that of the inspection of glass panels or face plates for the fronts of television picture tubes. The inspectors in the company in question originally accepted or rejected panels in terms of their memory of what constituted acceptable *vs.* unacceptable panels. The company then started using limit samples, one to represent each type of common defect. The company subsequently estimated that accuracy of inspection was increased 76 per cent after the inauguration of the new inspection procedure.

The procedures for the selection of limit samples in this instance consisted of first arranging a sample of several panels in order of "goodness." [20] This was actually done by the paired-comparison method, which was discussed in Chapter 9. This method, which involves the

[20] Martha L. Kelly, "A Study of Industrial Inspection by the Method of Paired Comparison," *Psychological Monographs*, 69 (9) (No. 394, 1955), 16 pp.

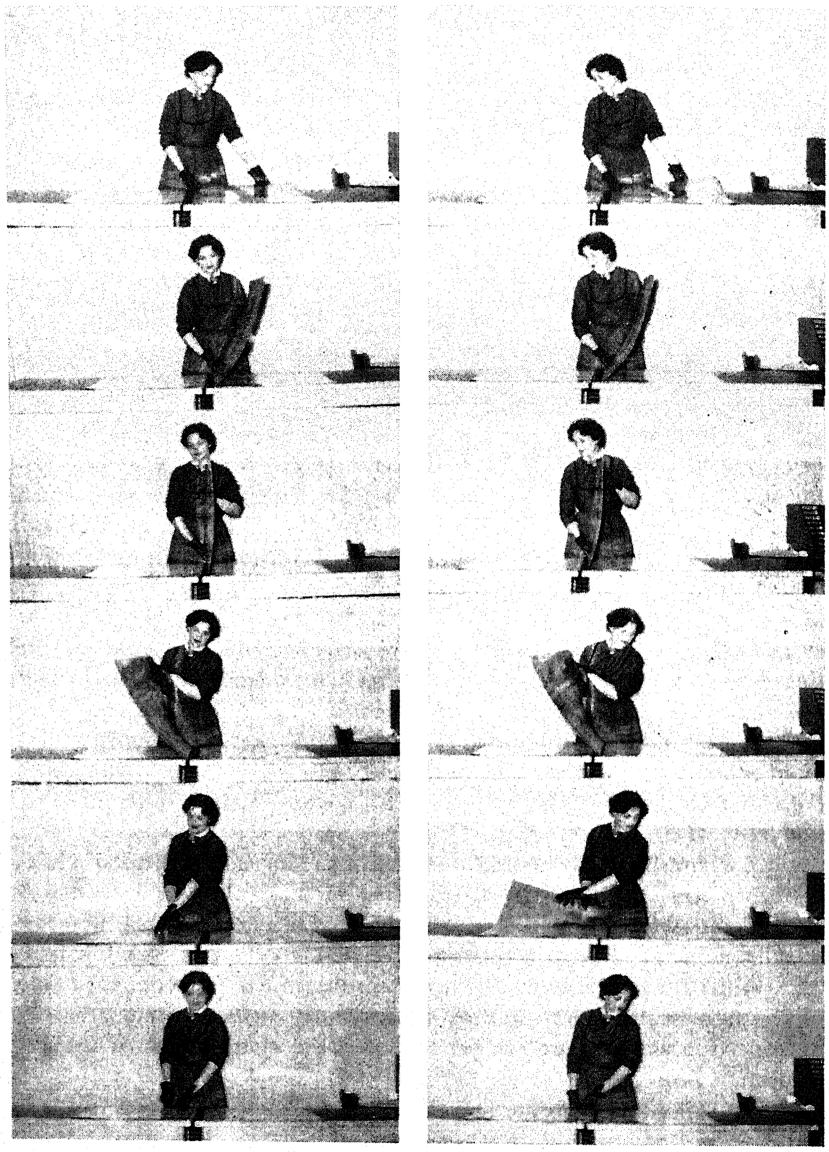


Fig. 17.8. For descriptive legend see opposite page.

judgment of each pair of objects, results in a scaling of the items. With a known "limit sample" panel in the lot to be scaled, one of the other panels was found to have nearly the same scale value as the true "limit sample" panel. This panel was then considered to be a "limit sample equivalent" and was given to an inspector to keep at her work place and to be used for comparison as additional panels were inspected. It should be mentioned that limit sample equivalents were identified in this way for each of several types of defects in order to have for each inspector a limit sample equivalent for each type of defect.

Layout of work area. In some inspection processes the layout of the work area, or the manner in which the products are displayed to the inspectors, can have an important influence on the quality of inspection. An example will illustrate this.

Where products to be inspected are moving, as on a conveyor belt, there seems to be some evidence to support the contention that they should move laterally *past* the inspector rather than move *toward* the inspector. The desirability of this practice was indicated by the results of a study by Kephart and Besnard.[21] In their experiment, clear glass spheres were inspected for appearance under two test conditions—with the spheres rolling toward the inspector and with the spheres rolling laterally past the inspector. Half of the spheres used had been scratched. It was the job of the inspector to identify these "bad" spheres. The results showed significantly greater accuracy with the lateral view than with the end view. Although this particular inspection job would probably not often be encountered in exactly the same form, the basic point emphasized by the authors is that the *method* of investigating accuracy that may be expected with different work layouts while equipment is still in the design stage will clearly contribute to the most effective final layout of work to be performed.

Method of display. An interesting method of investigation was that used by one company in connection with the job of inspecting roller bearings.[22] An eye camera was used to photograph the eye movements of 20 inspectors. On the basis of these films it was possible to determine

---

21 N. C. Kephart and G. G. Besnard, "Visual Differentiation of Moving Objects," *Journal of Applied Psychology*, 34 (1950), 50–53.

22 "Eye Camera Shows How to Improve Inspection Technique," *Factory Management and Maintenance*, 111 (No. 12, December, 1953), 88–89.

Fig. 17.8. Sequence of sheet movements and eye-movement patterns in inspecting tin plate. In the "Natural" way the eyes follow moving objects. The inspector thus sees the full sheet only in steps 1 and 5; she tries to see it while it is in motion in steps 2 and 4; she sees only the edge in step 3; and merely watches her hands in step 6. In the Right Way, which must be learned, the inspector sees a motionless sheet during all 6 steps. This way nearly doubles the available inspection time.

the extent to which the inspectors used *foveal* vision or *peripheral* vision. A person uses *foveal* vision when gazing straight ahead at an object; the object is then usually in clear focus. Objects outside this central area of visual attention are seen by *peripheral* vision.

The analysis showed that the portion of the surface of the roller bearings that was typically seen by foveal vision varied for inspectors of different ability levels, as follows: best inspectors, 88 per cent; average inspectors, 68 per cent; and poorest inspectors, 32 per cent. It was found, further, that visual inspection was optimum when approximately five-eighths of the total surface of the rollers could be seen by foveal vision, and the remainder by peripheral vision.

INDUSTRIAL INSPECTION

Information such as that obtained in this investigation was then used in designing the machines that rolled the rollers past the inspectors in order to provide for the use of foveal vision for about five-eighths of the rollers' surfaces.

In another inspection operation, that of inspecting clear glass bulbs, it was found that a striped background served as an aid to the inspection.[23] By viewing the glass bulbs against a striped background, with adequate illumination, it was found that defects in the glass caused the straight background stripes to appear wavy.

It should again be pointed out that examples such as these are given *not* so much for their *specific results, as to illustrate a method;* in particular they indicate how some systematic analysis of inspection operations (or, for that matter, of other jobs) can aid in determining what work method or work layout would be best for the job in question.

Selection of Inspectors From the above and other studies it is evident that the reduction of human errors in inspection frequently can be attacked by systematic attention to situational aspects (such as methods of inspection) and individual differences. While the consideration of individual differences in personnel selection was discussed in earlier chapters, one study dealing with the selection of inspectors will be illustrated here. In the study of bottle selectors by Coe,[24] an analysis was made of the relationship between certain items of biographical data and inspection ability as determined by a job sample test. Of 11 items of biographical information, it was found that 5 differentiated between "superior" and other selectors. The results, shown in Fig. 17.9, indicate that, generally speaking, the superior selectors tended to have the following character-

[23] "Lighting Up for the Inspector," *Modern Industry*, 22 (No. 6, December 15, 1951), 93.

[24] Coe, *op. cit.*

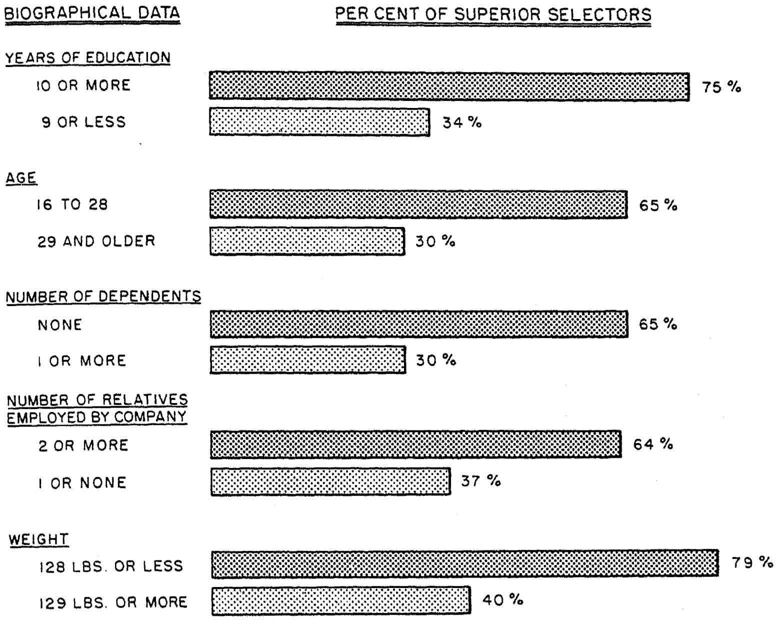


Fig. 17.9. Percentage of "superior" bottle selectors for various biographical data categories. (Adapted from Coe, *op. cit.*)

istics: they had had 10 or more years of education; were between 16 and 28 years of age; had no dependents; had two or more relatives who were also employed by the company; and weighed 128 pounds or less. These results could then be used as the basis for employing individuals who, on the average, might be expected to perform reasonably well on the job.

Training of Inspectors The training of inspectors (as the training for other jobs) needs to be specific to the job, and should be predicated on sound principles and practices. While training was discussed in Chapter 10, brief mention will be made here of the training of tin-plate inspectors mentioned earlier in this chapter. The inspectors were brought together in groups of 20, and were given the results of their own job sample tests, including information on their own accuracy in identifying each of the four types of defects. They were also shown examples of the different defects. To train them in the new method of inspection, they were shown micromotion films of the "normal" method of inspection and of the preferred method. This was accompaned by an explanation of eye movements with respect to the moving sheet. The films illustrated these eye-movement patterns and provided a means of teaching the most effective sequence of visual fixation. A general improvement in inspection performance followed these demonstrations.

18

Accidents and Safety

In the frame of reference of the previous chapter, accidents may be considered as errors. But—as in the case of other types of errors—accidents do not just happen. Rather, they can be considered as being brought about by certain preceding circumstances and events, some of which may be associated with human beings, others by situational factors (sometimes called mechanical factors).

Depending upon circumstance or purpose, accidents have been defined in various ways. Heinrich[1] defines an accident as an unplanned and uncontrolled event in which the action or reaction of an object, substance, person, or radiation results in personal injury or the probability thereof. He points out that, as in mathematics, there are theorems in accident occurrence, as, for example:[2]

1. A personal injury occurs as the result of an accident.
2. An accident occurs only as the result of a personal or mechanical hazard.

---

[1] H. W. Heinrich, *Industrial Accident Prevention*, 4th ed. (New York: McGraw-Hill Book Company, 1959).

[2] *Ibid.*, p. 17.

3. Personal and mechanical hazards exist only because of the faults of persons.
4. Faults of persons are inherited or acquired by environment.

Presumably those theorems would not apply to acts of God.

The implication of these theorems is that an injury cannot possibly occur unless there has been a personal unsafe act and/or exposure to an unsafe mechanical condition. Systematic efforts to minimize accidents would then seem to depend initially upon the availability of knowledge regarding unsafe acts and unsafe mechanical conditions. Such knowledge could then lead to appropriate corrective action.

THE HUMAN AND ECONOMIC COSTS OF ACCIDENTS

The number of fatalities from accidents is around 100,000 per year. These are broken down into classes as follows (data for 1963): [3]

All accidents	100,500
Motor vehicle	43,400 [4]
Public (excluding motor vehicle)	17,000
Home	29,000
Work	14,200

During that year there were about 2,000,000 disabling injuries in industry including about 80,000 cases of permanent impairment. On the job and off the job injuries totalled about 4,300,000. The National Safety Council [5] estimates economic loss as follows:

Wage loss, medical expense, overhead insurance costs	$2,650,000,000
Indirect costs (such as interference with schedules and property damage)	2,650,000,000
Total	$5,300,000,000

Such human and economic costs add up to staggering totals, and make it obvious that a major effort to lessen them is in order. Because of the magnitude of the situation, many companies have set up safety programs of various types to deal with the problem.

ACCIDENT AND INJURY DATA

The type of analysis that is required for a systematic attack on accidents requires adequate standardized data about accidents and injuries.

---

[3] *National Safety News,* 89 (No. 3, March, 1964), 46.

[4] Note: Motor vehicle total includes about 3,100 fatalities that are also included in the home and work totals.

[5] *National Safety News, op. cit.*

First, however, let us distinguish further between accidents and injuries. Heinrich [6] points out that the distinction between major or minor accidents is misleading In terms of his definition, there are many accidents that do not result in injuries—but the central point is that they *could* result in injuries. Thus, he argues that *accidents* are the point of attack (in taking corrective action) rather than *injuries* as such; thus the control of accidents should be directed toward the reduction of the unsafe acts and the elimination of the mechanical hazards.

Injury Data The American Standards Association [7] has proposed standard methods of recording accident records and of reporting accident statistics.

Injury records. Following the American Standards Practice, records of individual injuries are classified in terms of each of several accident factors. These are grouped in six major classifications as follows:

1. The agency (the object or substance most closely related to the injury, and which, in general, could have been properly guarded or corrected)
2. The agency part (the particular part of the agency that is most closely associated with the injury, such as grease, chuck, drill, etc.)
3. The unsafe mechanical or physical condition (the condition of the agency that could have been guarded or corrected)
4. The accident type (the manner of contact of the injured person with the object or substance, such as striking against, falling, slipping, etc.)
5. The unsafe act (such as making safety devices inoperative, failing to use protective goggles, etc.)
6. The unsafe personal factor (any mental or bodily characteristic which permits the unsafe act, such as lack of skill, bodily defects, etc.)

In addition, there is provision for recording the nature of the injury and duration of disablement (if any). In this connection a *disabling* injury is one which disables the injured worker for more than the day or shift on which he was injured. Disabling injuries are sometimes referred to as lost-time cases.

It should be added that in some accident statistics there is a distinction made for injuries that are not "disabling," as follows: (1) a *first-aid* case (this involves only first-aid care, but no loss of time since the worker returns immediately to his job); and (2) a *home-case* (in this case the worker returns home after the accident, but comes back to work the next day, losing no time beyond that lost on the shift on which he was injured).

---

[6] Heinrich, *op. cit*, p. 28.

[7] *American Standards Association,* American Standard Method of Recording and Measuring Work Injury Experience, Z16.1, 1954.

Injury statistics. The *injury frequency rate* is the number of disabling injuries per 1,000,000 man-hours, as reflected by the following formula:

$$\text{Injury frequency rate} = \frac{\text{No. of disabling injuries} \times 1{,}000{,}000}{\text{Total no of man-hours worked}}$$

The injury severity rate is defined as the number of days of lost time per 1,000,000 man-hours worked, as derived by the following formula:

$$\text{Injury severity rate} = \frac{\text{No. of days lost} \times 1{,}000{,}000}{\text{Total no. of man-hours worked}}$$

In case of death or permanent injury, the number of days lost is equated to 6,000, which is the average work days expectancy which such individuals would have had had they not been killed or permanently disabled.

In connection with severity of injuries, Heinrich refers to a 300-29-1 ratio, as follows:

For every	300	no-injury accidents (90.0%)
There are	29	minor injuries (8 8%)
and	1	major lost-time (disabling) injury (0.3%)

GENERAL CONSIDERATIONS

As implied above, it is the objective of accident and injury research to trace the factors that contribute to unsafe acts and to unsafe mechanical conditions, in order that appropriate action can be taken based on that knowledge. This is, in effect, a distinction between personal (psychological) variables and situational variables.

"Causes" of Accidents In the investigation of individual accidents, it is usually the practice to try to identify the specific cause (or causes) of the accident in terms of specific unsafe acts, or unsafe mechanical or situational conditions. It might be found, for example, that many accidents are "caused" by removal of guards on the machines or by improper work methods. As pointed out by Heinrich,[8] cause analysis is a necessary prerequisite to the selection of an effective remedy. While such "causes" are the immediate ones, accident causation sometimes can be characterized at other levels of abstraction.

Analysis of Relationships Aside from causation as such, an analysis of accident statistics frequently can reveal relationships with relevant variables that add understanding, and that can lead to appropriate corrective actions. For example, Table 18.1 shows the rates for lost-time and first-aid injuries of personnel in 11 departments of a steel mill. As expected, the rates for first-aid injuries were higher than for lost-time

[8] Heinrich, *op. cit.*, p. 101.

injuries, but the more relevant aspect of these data is that the departments that had high rates for lost-time injuries did not necessarily have high rates for first-aid cases. For example, Department 5, which is highest in terms of lost-time cases with a rate of 8.37, is seventh from the top in first-aid cases with a rate of 496. The rank order coefficient of correlation between these two sets of rates was only .21. In this situation, then, it appears that the factors which result in a large number of first-aid injuries (whatever those factors may be) are not necessarily the same as those which result in a correspondingly large number of lost-time injuries.

Table 18.1

LOST-TIME AND FIRST-AID INJURIES FOR ELEVEN DEPARTMENTS OF A STEEL MILL FOR ONE FISCAL YEAR

Department	*Lost-time injuries* *Rate per million hours*	*Rank*	*First-aid injuries* *Rate per million hours*	*Rank*
1	5.91	2	433	9
2	3.30	4	660	3
3	2.36	8	574	4
4	2.34	9	485	8
5	8.37	1	496	7
6	2.62	6	392	10
7	4.18	3	930	1
8	2.45	7	725	2
9	2.65	5	504	6
10	2.26	10	534	5
11	.50	11	242	11

One should be cautious, however, in generalizing from a finding of this sort. In other situations, particularly in other industries, a more marked relationship between the frequency of minor and major accidents may be found. Heinrich,[9] for example, points out that severity of injury can be variable within the same type of circumstance.

Aside from such analyses as the relationship between injuries of varying degrees of severity, accident and injury statistics can be analyzed in particular with the view toward determining the extent to which accident or injury frequency is associated more with *individuals* or with *situations.* Further, in the case of significant individual differences, it is sometimes possible to identify by statistical analyses what personal characteristics are associated with accident occurrence. It is to these matters that we will now turn our attention.

Accident Proneness When industrial safety first began to be given serious consideration by management, opinion differed considerably on the relative importance of *factors within the individual* and the *work*

[9] Heinrich, *op. cit.*, pp. 28–29.

situation as causes of accidents. It is obvious that a safety program within a plant must depend upon specific knowledge of whether these factors are both operating, and, if they are, what their relative importance is for the plant in question. The concept of "accident proneness" is concerned with whether each employee tends to retain a given accident rate in comparison with other employees regardless of changes in the kind of job or the work situation. To the extent that each employee does tend to retain his same relative accident rate from one time to another, we may conclude that personal factors, peculiar to that individual, are affecting his accident record. This situation has been called *accident proneness.* As used in this text, accident proneness is a descriptive rather than an explanatory term. When an individual employee is found consistently to experience more accidents than the average employee, he may properly be classified as an accident-prone employee. Such classification does not, however, explain *why* he is accident-prone; an explanation would require identification of the *causes* of his accident proneness.

It should be noticed in the above discussion that we have identified the accident-prone employee as one who *consistently* has more accidents than the average employee. An error has sometimes been made in accident studies when those employees who, during *one* period of time, have more accidents than the plant or department average are singled out as accident-prone employees. Mintz and Blum [10] have clearly shown that if only chance factors are causing the accidents, employees will not all be equally lucky or unlucky during a given period of work. If chance factors alone were operating, the situation would be quite analogous to tossing coins. If 1,000 men were each to toss ten coins, the average number of heads per man would be very close to five. But the individual men would vary in number of heads from very few (perhaps none) to many (perhaps ten). If we simply identify those men who tossed heads eight or more times, and call these men "heads-prone," we would be making the same error that is made when men who have more than the average number of accidents for one period of time are called accident-prone. But if such men are identified from accident records for a specified period of time and during a subsequent temporal period again experience more accidents than the average, we are reasonably safe in assuming that accident-proneness factors are operative. Maritz [11] has pointed out that the technique of correlating records for consecutive periods of time would seem to be indispensable in demonstrating the presence of accident-proneness factors.

---

10 A. Mintz and M. L. Blum, "A Re-examination of the Accident Proneness Concept," *Journal of Applied Psychology*, 33 (1949), 195–211.

11 J. S. Maritz, "On the Validity of Inference Drawn From the Fitting of Poisson and Negative Binomial Distributions to Observed Accident Data," *Psychological Bulletin*, 47 (1950), 434–443.

The relative importance of working situations and accident proneness as causative factors varies from one plant to another. In plants where the machinery used is unavoidably dangerous in itself, the former is of greater importance. In plants where the working conditions in and of themselves are not hazardous, the latter factor is of greater importance. When individual or personal factors are causing the trouble, it is clear that an adequate safety program must make provision for the individual handling of the high-accident-rate or accident-prone employees. It is, therefore, of vital importance for the safety director of a plant to know the extent to which his own employees are affected by accident-proneness factors.

In nearly every investigation of industrial accidents, accident proneness has been found to be a factor, and, in some cases, a vitally important factor. The early work of Greenwood and Woods,[12] Marbe,[13] and others tends to support the statement that accidents happen frequently to some men and infrequently to others as a logical result of a combination of circumstances. Several more recent investigations substantiate this general conclusion that, at least in many jobs, some employees tend to have consistently more accidents than can reasonably be accounted for by chance. In a study carried out by one of the authors, for example, an injury index was derived for employees on three jobs, this index being the frequency per 1,000 hours worked. This index was derived separately for two six-month periods, and these were then correlated. The correlations are given below:

Job	*Correlation*
Drill press	.86
Assembler	.88
Machine operator	.74

Mintz[14] has also reported that accident records of individuals have some validity as indications of future accident rates of at least some of those individuals, particularly those in which the persons have had a large number of accidents. It is not our intention here to summarize the literature on this subject. Our purpose, rather, is to present certain statistics that have a direct bearing on this topic. These statistics both strengthen the case for the concept of accident proneness in industrial accidents and answer certain questions that previous investigations have not particularly considered.

---

12 M. Greenwood and H. M. Woods, "The Incidence of Industrial Accidents, with Special Reference to Multiple Accidents," *Industrial Fatigue Research Board, Report No. 4* (London, 1919).

13 K. Marbe, *Praktische Psychologie der Unfalle und Betriebschaden* (München, 1926).

14 A. Mintz, "The Inference of Accident Liability from the Accident Record," *Journal of Applied Psychology*, 38 (1954), 41–46.

Table 18.2 summarizes certain facts pertaining to hospital visits from the employees of 11 departments of a sheet and tin mill employing around 9,000 employees for two years. The second, third, and fourth columns show, respectively, the average number of hospital visits per man for the first year, for the second year, and for the two years combined, for each of the 11 departments as well as for the mill as a whole (in the bottom row). In the remaining columns of this table the average hospital visits per man for the second year are given for the employees who had previously been classified according to the number of hospital visits experienced in the first year. Thus, for Department 1, the individuals with no hospital visits during the first year averaged .61 hospital visits in the second year, whereas in this same department the employees with seven visits in the first year averaged 2.50 visits in the second year. Although some inversions of the data from one category to the next occur, particularly in departments where the total number of men was not large, the general trend in all departments is toward an increase from left to right. When attention is turned to the data summarizing the mill as a whole, which are tabulated in the lower row, this trend is seen to be still more pronounced. For the mill as a whole, no inversions whatever are found in the tendency.

These data for the entire mill are shown graphically in Fig. 18.1. In this figure the number of hospital visits for the first year is plotted along the base line or abscissa and the corresponding number of hospital visits for the following year along the ordinate. From these data we may confidently state that the average employee in this mill tends to retain approximately his same relative position with regard to hospital visits for any two successive years.

Accident Proneness *vs.* the Job in Relation to Accidents The question may be raised why those employees who had nine hospital visits during the first year did not, on the average, have nine such visits during the second year instead of 5.14 visits as revealed by the data. If factors within the individual (accident proneness) were the sole cause of hospital visits, then the employees with nine hospital visits during the first year should also have had on the average nine such visits during the following year. The discrepancy between the nine visits experienced by employees in this group during the first year and the 5.14 average visits experienced by these same employees in the following year may be taken as indicative of the relative relation of accident-proneness factors *vs.* external factors to hospital visits in the mill in question. In other words, we may say roughly that of every nine hospital visits made by the most accident-prone group, slightly more than five are essentially related to the accident proneness of these individuals and the remaining four cannot be related so clearly to individual factors within the employees. To prove the importance of the accident-proneness hypothesis, it is not necessary that the high-accident

Table 18.2

FREQUENCY OF HOSPITAL VISITS FOR TWO CONSECUTIVE YEARS IN ELEVEN DEPARTMENTS OF A STEEL MILL

Department	Freq. 1st year	Freq. 2nd year	Freq. both years	Average number of hospital visits in second year for men with 0, 1, 2, etc., visits in first year — Number of visits in first year: 0	1	2	3	4	5	6	7	8	9
1	66	.85	.75	.61	1 05	1.54	1.14	1.64	2.20	5.00	2 50		
2	1.13	1.31	1.22	.68	1.31	1.60	2.35	3.63	3.27	3.00	4.33		3.50
3	.81	1.03	.96	.63	1.14	1.56	2.16	2.56	4.50	3.00	1.00		7.00
4	.89	.81	.94	.55	.87	1.06	1.54	1.86	1.78	2.66	4 25	5 00	
5	1.01	.99	1.00	.68	.95	1.00	1.94	2.23	1.60	3 33	5.00		
6	1.20	1.27	1.23	.74	1.24	1.56	1.98	1.82	3.63	3.63	5.00		7.00
7	1 12	1 26	1.19	.81	1.15	1.45	2.00	2.29	3.44	3.21	11.00	4.00	5.00
8	1.32	1.21	1.26	.64	1.04	1.35	2.34	2.28	2.61	3.36	4.00	6.50	4.70
9	.71	1.05	.88	1 03	1.11	1.43	1.48	3.15	4.50	7.00			
10	.91	1.16	1.03	.53	1.26	2.05	2.41	3.18	3.50	4.25	4.50	7.00	8.00
11	.58	.88	.73	.71	1.22	1.47	1.14	3.00	5.50				
Average for all depts.	.94	1.07	1.01	.69	1 12	1.46	1.86	1.91	3 32	3.84	4 62	4.91	5.14

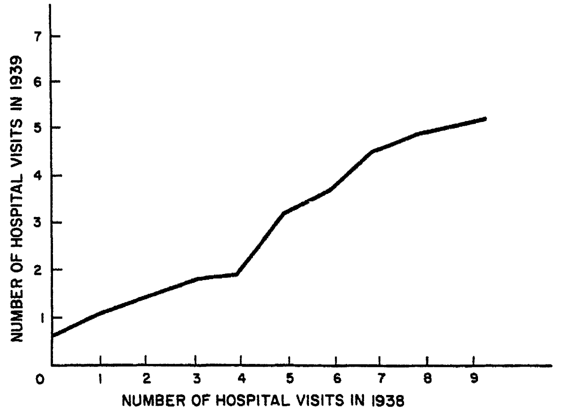


Fig. 18.1. Relation between number of hospital visits for two successive years among approximately 9,000 steel workers.

group for any year retain the same level of accidents for a subsequent year, but only that this group account during the subsequent year for a number of accidents significantly greater than the average of the plant. As revealed in Table 18.2, this situation is clearly the case for the mill in which these statistics were gathered.

An objection may be raised to these results as evidence of the fact that an individual or personal factor operates in the causation of accidents. This objection is that the job to which each employee is assigned

Table 18.3

VARIATION IN AVERAGE NUMBER OF HOSPITAL VISITS PER YEAR AMONG EMPLOYEES ON ELEVEN DIFFERENT JOBS

Job	*Average number of hospital visits per year*
Craneman	3.55
Opener	3.54
Reckoner	2.96
Machinist helper	2.77
Leader	2.75
Sheet inspector	2.54
Shear helper	2.40
Assorter	2.36
Potman	2.10
Foreman	1.16
Roll turner	.47

may involve a specific accident hazard that, remaining the same or relatively the same from one year to another, results in a consistency of accidents for each employee from year to year, a consistency due more to the accident hazard of a given job than to any factors within the individual. It must indeed be admitted that each job has a relatively constant accident hazard and that this hazard clearly varies by rather marked amounts from one job to another. This fact is brought out in Table 18.3, which gives the average number of hospital visits per year for employees on 11 specific jobs in a steel mill. It will be noted that the average number of hospital visits varies from 3.55 in the case of cranemen down to .47 in the case of roll turners. It might reasonably be expected that this ratio of 7 to 1 found among this sampling of 11 jobs would be still larger if a larger sampling of jobs had been included in this survey. The table will suffice, however, to set forth clearly the fact that a specific hazard on different jobs does exist and that this hazard is likely to vary by a rather large amount from one job to another.

Table 18.4

SIGNIFICANT † CORRELATIONS BETWEEN PERSONNEL OR DEPARTMENTAL DATA AND ACCIDENT FREQUENCY AND SEVERITY

	Accident frequency	*Accident severity*
Per cent of employees who are male, production		63
Per cent of employees who are male, salary		.50
Intra-company transfer mobility	−.44	
Sex ratio imbalance (men to women)		51
Mean promotion possibility (mean ratings, twelve judges)	−.40	−.50
Per cent of suggestion quota met		−.54
Mean noise level	.42	
Youthfulness of employees (per cent under 26)		−.57
Tenure (per cent employed more than 12 months)		.55

Source. *Kerr,* op cit.
† *Less than one chance in twenty that they could have occurred by chance.*

The fact that the accident hazard varies significantly among different departments of a plant is further revealed by a study made by Kerr.[15] His study dealt with 53 departments of the Camden Works of the Radio Corporation of America. Correlations were determined between both injury frequency and injury severity with 40 types of personnel or departmental data. The correlations found to be significant are given in

[15] W. A. Kerr, "Accident Proneness of Factory Departments," *Journal of Applied Psychology*, 34 (1950), 167–170.

Table 18.4. Kerr suggests that a common explanatory factor of the relationships found might be *depressants to alertness.* But whether this is or is not a satisfactory explanation of the findings, the fact remains that significant relationships were found between injury rates and certain characteristics of the departments.

This situation makes it necessary to rule out the element of job differences in evaluating the concept of accident proneness. This was done by reassorting the data from the steel mill according to specific jobs. Thus, all openers, regardless of the department in which they are working, are exposed to approximately the same accident hazards. The combined data for the 11 jobs considered are portrayed graphically in Fig. 18.2. We note

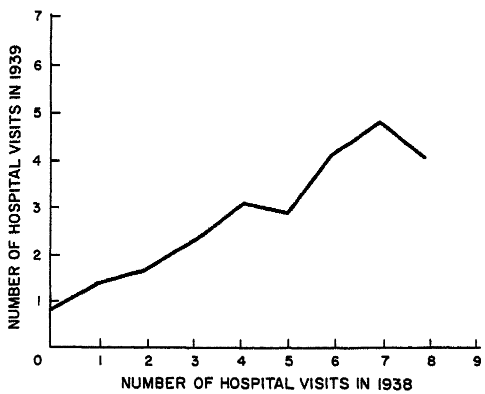


Fig. 18.2. Relation between number of hospital visits for two successive years among steel workers when hazard of the job is eliminated as a determining factor.

nearly as definite a relationship between hospital visits of two successive years in Fig. 18.2, when the element of job hazard has been essentially eliminated, as we previously noted in Fig. 18.1, when the job-hazard element might conceivably have had a significant effect upon the results. From the data, it seems safe to conclude that accident proneness, or individual accident susceptibility, is significantly related to the type of accident that results in hospital visits for the plant studied.

Further insight into the matter of accident proneness is furnished in an article by Whitlock, Clouse, and Spencer.[16] These men, having

[16] G. W. Whitlock, R. J. Clouse, and W. F. Spencer, "Predicting Accident Proneness," *Personnel Psychology*, 16 (1963), 35–44

noted that there is typically a rather low correlation between number of injuries in successive periods of time, were concerned in their investigation with what they called "accident behavior." They defined accident behavior as that which might result in (a) no injury, (b) an injury to the person committing the behavior, (c) an injury to someone else, or (d) both b and c above. This is essentially the same as the definition of accidents by Heinrich, mentioned earlier.[17] Examples of accident behavior (i.e., unsafe performance) were obtained from first-line supervisors. After combining similar examples and discarding those which were exceedingly rare, there remained 76 examples of unsafe behaviors. Thirty first-line supervisors then made daily records of the occurrence of such unsafe behaviors during a period of 8 months.

After some additional preliminary investigation, the format for recording unsafe behaviors was improved and a further study was conducted. In this investigation, the reliability of the unsafe behaviors record was found to be .93. Data were also obtained at this stage on actual injuries sustained by each employee during a six-year exposure period. The correlation between unsafe behaviors observed during an 8-month period and actual injuries during a 6-year period was found to be .56. This investigation strongly supports the idea that unsafe behaviors are something that some employees consistently practice. The investigation further shows that employees who consistently behave unsafely sustain a larger proportion of the actual injuries than do employees who consistently follow safe practices.

It is one thing to determine, as the data above demonstrate, that an appreciable number of injuries are due to accident proneness or factors within the individual. It is quite another thing, however, to identify the particular factors within the individual that determine accident proneness. In other words, once the concept of accident proneness has been established as an operating factor, the next question that arises is: For a particular job, what are the specific personal factors which differentiate between the consistent accident repeaters and employees with more favorable accident records? A discussion of this problem follows later in this chapter.

SITUATIONAL FACTORS RELATED TO ACCIDENT EXPERIENCE

There are many different kinds of situational factors that can—in different circumstances—be related to accident experience. The nature of the activity being carried out, for example, can influence accident frequency; the data in Table 18.3 on hospital visits for employees on

[17] Heinrich, *op. cit.*

different jobs illustrate this. Crawford[18] has also emphasized the importance of the job itself in relation to accident frequency. He reported that third class linemen had an accident rate of .65 accidents per employee, which was significantly higher than the rate of .30 per employee for all employees in the division studied.

Closely related to differences in hazards associated with jobs are differences associated with departments or work units. Table 18.5, for example, shows the hospital visits for ten departments of a sheet and tin mill. It will be noted in the last column of this table that the hospital visits per man per year vary from a minimum of .55 to a maximum of 1.26. This difference clearly reflects the hazard of working in the different departments.

Table 18.5

DEPARTMENTAL DIFFERENCES IN HOSPITAL VISITS RATE

Department	*Number of men*	*Number of accidents during two years*	*Accidents per man per year*
1	811	902	.55
2	573	1,144	1 26
3	480	723	.75
4	1,099	1,599	.73
5	336	555	.67
6	582	1,123	.96
7	624	1,238	.98
8	1,192	2,266	.94
9	373	529	.65
10	1,101	1,945	.88

Depending upon the circumstance, other factors besides the job and department can influence the accident rate. These include the design of the equipment used, the physical condition of the equipment, whether guards are installed and used on equipment, work methods, housekeeping practices, illumination and other working conditions, work schedules, and time of day. In connection with work schedules, for example, it has been pointed out by Vernon[19] that the accident rate increases during the latter part of the working day. According to Vernon's results, this tendency is so marked that during a 12-hour working day women experienced two and one-half times as many accidents as during a 10-hour day. Although fatigue has often been considered the cause of this increase, the fact that the time of greatest accident rate as compared with hours worked

---

[18] P. L. Crawford, "Hazard Exposure Differentiation Necessary for the Identification of the Accident-Prone Employee," *Journal of Applied Psychology*, 44 (1960), 192–194.

[19] H. M. Vernon, "An Experience of Munition Factories During the Great War," *Occupational Psychology*, 14 (1940), 1–14.

is reversed on the night shift indicates that psychological rather than physiological factors are operating. Although Vernon's results attach somewhat more importance to the length of the working day than do other investigations of this subject, it is quite commonly agreed that as the working day is lengthened, the accident rate increases in greater proportion than the increase in number of hours worked.

It is also quite commonly accepted among safety men that an increase in the accident rate usually accompanies a step-up in production. Although this situation undoubtedly occurs in many individual plants when an attempt is made to step up production without adequate expansion of plant facilities, it is interesting to note that over a period of years, the rate of injuries has gone down very markedly. Blake,[20] for example, estimates that had the injury rates of about 40 to 50 years ago been continued until today, there would now be at least two and one-half times more injuries than there actually are. This means that there is nothing inherently characteristic of a high production rate that tends to make automatically for a high accident rate. Rather it means that, as high production is achieved in the normal course of events by improved machinery and capacity for output, the accident rate will be likely to decrease along with the technological improvements and plant expansion.

Kerr and his co-workers have published several interesting studies which reveal relationships between accident experience and a variety of what we might call social situational factors.[21] In one of the studies, injury *severity* and *frequency* were separately correlated with 75 other variables. Data were obtained from 147 plants in the automotive and machine shop industry.

Injury frequency was found to be greatest in plants with a high seasonal layoff rate, where employees looked with disfavor on high-producing coworkers, where there were many other similar plants in the neighborhood, where employees frequently needed to lift heavy materials, where there were blighted living conditions, and where there was a record of many garnisheed wages.

Injury severity was found to be greatest where employees and management personnel did not use the same dining room, where national unions were very strong, where there was no stated penalty for tardiness, where there was no employee profit-sharing plan, where there were extreme work place temperatures, and where the work was "dirty-sweaty."

Many of the findings seem to show in common the operation of a persistent threat to, or undermining of, the status or comfort of the individual employee (seasonal layoffs, rival hostility among employees, social distance in eating practices, dominance of a national union in

---

[20] R. P. Blake, *Industrial Safety*, 3rd ed. (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1963).

[21] P. Slivnick, W. Kerr, and W. Kosinar, "A Study of Accidents in 147 Factories," *Personnel Psychology*, 10 (1957), 43–51.

collective bargaining, no incentive in profit sharing, heavy and dirty work). Kerr feels that the loss of, or threat to, individuality may produce preoccupation which results in unsafe behavior.

Another group of correlates of high-injury plants relates to urban sociology (congestion, ugliness of other near-by plants, living conditions in blighted neighborhoods, and garnisheed wages). Kerr hesitates to attach any causal significance to these findings but he emphasizes that they should be kept in mind as factors that are associated with both *frequency* and *severity* of injuries in industrial plants.

As still another illustration of the differential accident liability of various situations, the Automotive Safety Foundation reports data on accident rates on various classes of highways, as given in Table 18.6. It can be seen that the accident rate per million vehicle miles is four times as high on four-lane undivided highways as on freeways.

Table 18.6

ACCIDENT RATES ON VARIOUS TYPES OF HIGHWAYS

Type of highway	*Accidents per million vehicle miles*
2-lane	2.38
3-lane	2.57
4-lane, undivided	4.09
4-lane, divided	2.91
Divided, controlled access	1.69
Freeway	1.00
Total	2.15

Taken from American Safety Foundation, "Traffic Control and Roadway Elements: Their Relationship to Highway Safety," Traffic Safety, *7 (No 3, September, 1963), 9–18, Table III.*

PSYCHOLOGICAL FACTORS RELATED TO ACCIDENT EXPERIENCE

It has been implied in the above discussion that accidents are caused, and do not simply happen. Basically, accidents probably can be attributed to either the *work situation* or to *personal variables* (accident proneness). The work situation generally determines to some degree the *liability* of accidents (some work situations are more hazardous than others). If, for people in the same work situation (and therefore subject to the same liability), there are consistent differences in accident frequency *among the individuals,* it can then be assumed that there are one or more *personal variables* that are contributing to the consistent individual differences. Knowing this, it is *then* in order to try to identify *what* personal variables are associated with accident frequency on the job in question. These will vary from job to job, and only through systematic investigation is it

possible to identify *what* variable or variables are operating in the particular situation at hand. A few of these variables will be discussed.

With respect to the influence of personal variables in accident causation, Heinrich [22] estimates that about 88 per cent of all accidents are caused primarily by the unsafe acts of persons. These "unsafe acts," of course, basically could be attributable to one or more human variables.

Before discussing some of these variables, however, it should be added that an analysis of the *accident liability of the work situation,* and of any *personal variables,* usually will not account for *all* of the variability in accident frequencies among individuals. That which we cannot explain, we attribute, then, to "chance." It should not be assumed, however, that if we cannot account for all accidents, the ones unaccounted for are really due to chance. Rather, we should simply assume that we have *not yet accounted* for those factors that *are* causing them.

Generalized Accident Tendencies It has been suggested by some investigators that there are certain pervading personal characteristics that, in the case of some individuals, cause them *generally* to have more accidents than others, in *whatever* situation they may be. If this is a valid postulate, such an individual could well be characterized in facetious terms as "an accident looking for a place to happen." Various types of characteristics have been suggested as the possible predisposing factors or as symptoms thereof, such as emotional and personality variables, motivation, aggressiveness, unwillingness to conform, willingness to take chances, above-average drive, low tolerance of frustration, strong guilt feelings, poor social adjustment, and generally inappropriate work habits.[23]

This general hypothesis gives rise to the use of the term accident proneness to characterize the underlying human characteristic that might be responsible for such a generalized accident tendency. Let us hasten to repeat, however, that in this text the term accident proneness is being used in the statistical (descriptive) sense discussed earlier, and not in an explanatory sense.

The hypothesis of a generalized accident tendency in the case of some people seems to be a reasonable one, and may very possibly be substantiated. In one industrial organization it was found that employees in a high-injury-frequency group tended, in general, to be undesirable employees from other viewpoints as well, such as having had more reprimands and more leaves (other than vacation) than did a low-injury-frequency group, thus suggesting something of a general pattern of lack

---

[22] Heinrich, *Industrial Accident Prevention,* 4th ed. (New York: McGraw-Hill Book Company, 1959).

[23] Some of these factors are discussed in such sources as *The Human Element in Industrial Accident Prevention* (Center for Safety Education, New York University, 1955), and *Accident Prevention Manual for Industrial Operations,* 4th ed. (National Safety Council, 1959), Chapter 5, "Human Behavior and Industrial Safety."

of adjustment of which accident frequency might be simply one manifestation.

At the present time, however, there probably is insufficient evidence to be able to support the notion of a generalized accident tendency. Under the circumstances, we should consider accident proneness (in its statistical sense) as being *situational.* Where, *within* a work situation with a given accident liability rate, we find statistically significant evidence of differential accident rates on the part of different individuals, we should then seek to identify what *particular* human characteristics are related to such differences. Some such characteristics will now be discussed. Before doing so, however, two more points should be made. In the first place, if in the future it is demonstrated that there is some generalized accident tendency, it is probable that it will *not* account for a large proportion of accidents. And in the second place, we need fully to recognize the possible influence on people—and their "accident behavior"—of personality, emotional, motivational and related variables in *individual* accident occurrence, and possibly as personal characteristics that may *in given situations* be generally relevant to accident frequency.

Vision In many different types of jobs, it has been found that vision is related to accident frequency. This problem has been investigated by comparing the accident experience of employees whose visual skills meet certain statistically determined standards with the corresponding accident experience of employees whose vision does not meet these standards. Table 18.7 summarizes data from several plants showing that employees whose

Table 18.7

RESULTS OF VISUAL SKILL TESTS IN RELATION TO INJURIES

		Percentage of employees who had not over two injuries during the period of the investigation among those who:	
N	*Job*	*Pass visual requirements*	*Fail visual requirements*
59	Passenger car drivers	71	42
116	Mobile equipment operators	74	71
65	Machine operators	67	50
15	Machine operators	67	44
29	Sheet metal workers	58	41
105	Maintenance men	81	75
63	City bus drivers	44	33
66	Intercity bus drivers	54	30
68	Supervisors	65	57
125	Machine operators	56	32
102	Laborers	82	89
15	Skilled tradesmen	45	33
828	All groups combined	65	57

vision is adequate for their jobs have fewer accidents than do employees with less adequate vision.[24]

For the 59 passenger car drivers in the first job group, 35 had visual skills that met the recommended requirements. Of these 35, 25 or 71 per cent showed a low frequency of accidents. On the other hand, among those passenger car drivers whose visual skills did not meet the standard requirements, only 42 per cent showed a low frequency of accidents. A higher percentage of those workers whose vision was adequate for the job were in the low-accident-frequency group. Table 18.7 shows that for each of the twelve job groups similar results were obtained for all groups except laborers. In eleven of the twelve job groups, a higher percentage of workers whose visual skills meet the requirements for their jobs were in the low-frequency-of-accidents category than the corresponding percentage in this category among those whose visual skills did not satisfy the job requirements.

Of the total group of 828 workers, 463 or 56 per cent (not shown in Table 18.7) had visual skills that met the requirements of their job standard. Of these 463 workers possessing adequate visual skills, 65 per cent were in the low-accident-frequency group. Three hundred sixty-five or 44 per cent of the workers did not have the visual skills required by the job pattern. Of these 365 only 57 per cent were in the low-frequency-of-accidents group. This is a difference of 8 per cent in favor of that group whose visual skills met the requirements of the standard. Statistical analysis (beyond the scope of this book) shows that the odds of these results occurring by chance are less than one in a hundred.

The results of one study of this sort are shown in Fig. 18.3. The visual standard for the job of paper machine operator was the one for machine operators as described on page 181 and illustrated in Fig. 6.16 in Chapter 6.

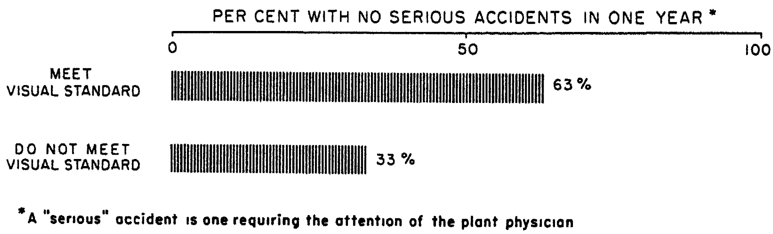


Fig. 18.3. Vision in relation to serious accidents among 104 paper machine operators. (Adapted from S. E. Wirt and H. H. Leedke, "Skillful Eyes Prevent Accidents," Annual News Letter, Industrial Nursing Section, National Safety Council, November 1945.)

[24] N. C. Kephart and J. Tiffin, "Vision and Accident Experience," *National Safety News*, 62 (1950), 90–91.

The results of these several studies clearly reveal the relationship between vision and accident experience on some types of jobs. When these statistics were first shown to safety men from one of the plants in question, a thorough study was made of the individual case records of the various lost-time accidents. Nowhere in the description of these accidents, which had been written at the time of the accident by persons familiar with the background of each mishap, could there be found reference to the fact that faulty vision had played a part in causing the accident. These individual case studies did not reveal a deficiency of vision of the employee responsible for the accident. Yet the statistics show that the visual factor was operating. On an actuarial basis, it can be predicted that, if all plant employees had been required to pass the critical vision tests, the lost-time accident rate among the employees would have been markedly decreased. Uncovering by statistical investigation the importance of this factor, which had long escaped notice through ordinary case study of accidents, is a striking illustration of the value of the statistical approach in revealing hidden factors in any phase of human behavior.

No one should conclude from these statistics, however, that faulty vision is the *sole* cause of industrial accidents. Just as all kinds of factors may cause disease, so all kinds of factors may cause accidents. But the statistics do prove that faulty vision is *one* of the factors related to accidents in some circumstances.

Age and Plant Service A number of the investigations of the relationship between accident rates and plant experience or job experience of the employee have been reported. As might be expected, these investigations do not reveal entirely consistent results from one plant to another or from one industry to another. In most of these studies it has been found that the accident rate is higher among the younger, more inexperienced employees. In one study carried out in Sweden, for example, Zetterman[25] shows a noticeable drop in disabling injuries with age, but no marked relationship between age and fatalities and permanent disability. These relationships are shown in Figure 18 4. On the basis of an analysis of injuries in a copper plant, VanZelst[26] also concludes that accident rates tend to be higher among younger persons. Data obtained in a sheet and tin mill also show a dropoff in the accident rate with the age of the employees, their years of service in the plant, and their years of service on their present jobs. These relationships are shown graphically in Fig. 18.5. Except for a low point of the curves for the very young or very inexperienced employees, there is a continuous decrease in hospital visits with increasing age or increasing plant or job experience.

---

[25] N. Zetterman, "Industrial Accidents—a Burden on the National Economy," *Occupational Safety and Health,* 1 (No. 2, April–June, 1951), 61–62.

[26] R. H. VanZelst, "The Effect of Age and Experience Upon Accident Rate," *Journal of Applied Psychology,* 38 (1954), 313–317.

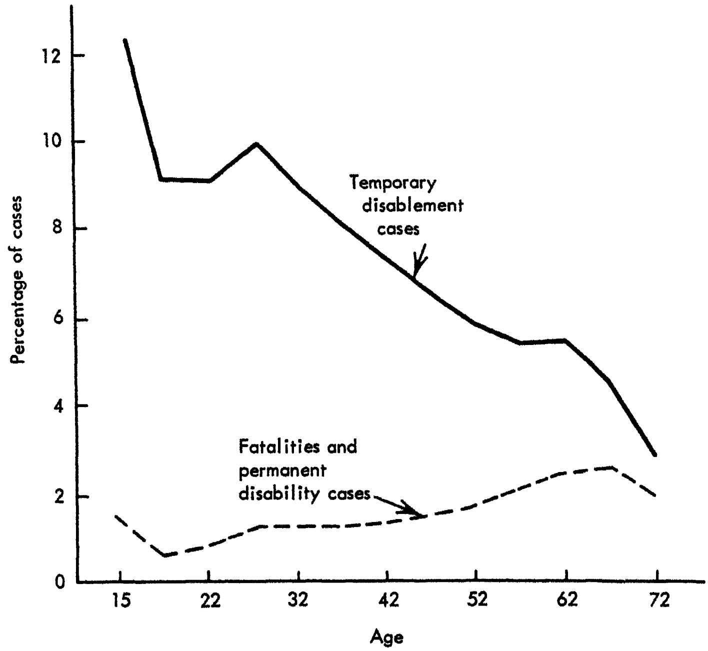


Fig. 18.4. Relationship between age and two classes of injuries (temporary disability versus fatalities and permanent disabilities). (From Zetterman, *op cit.*)

Several possible explanations for the trends shown in Fig. 18.5 have been offered. It has been suggested that the younger employees are placed on the more hazardous jobs and that, as they become older and more "plant wise," they tend to be transferred to jobs of greater relative safety. It has also been suggested that the younger employee, who, in general, has relatively few family responsibilities, is less cautious and more likely to take chances than the older employee and therefore becomes involved in a greater proportion of industrial accidents. VanZelst [27] refers to the immaturity of younger people. Whatever the underlying explanation, however, it is important to the successful operation of a safety program to know what relationships there are between such factors as age and experience on the one hand and accidents on the other.

It should be emphasized, however, that the relationships indicated above are not necessarily applicable in all situations. The important thing to remember is that no single rule or set of relationships necessarily applies on an across-the-board basis, and that such information should

[27] VanZelst, *op. cit.*

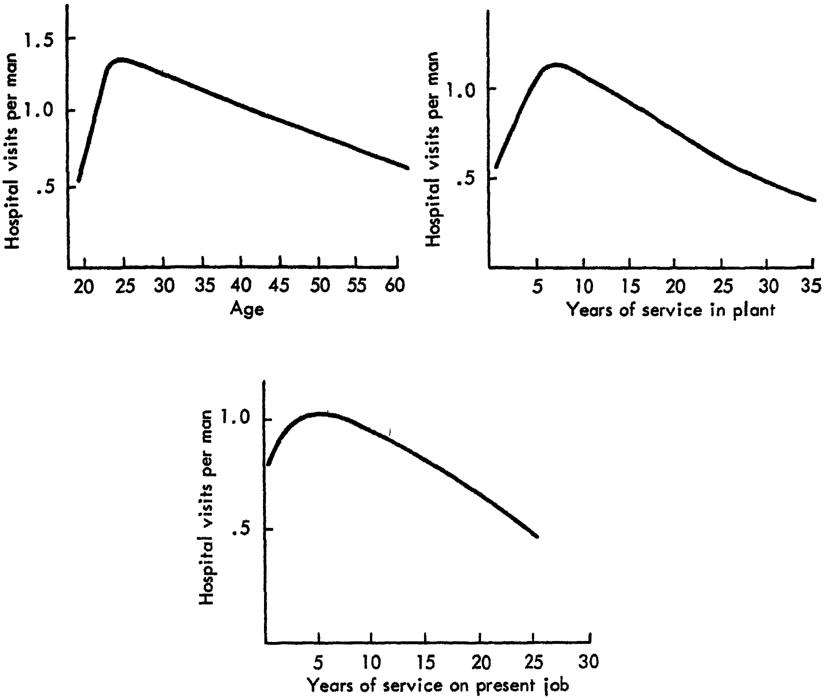


Fig. 18.5. Hospital visits per man per year in relation to age, years of service in the plant, and years of service on present job among 9,000 steel workers.

be determined for individual situations in order to plan the safety program with this information in mind.

Emotional Factors Two emotional factors that have been found to be related to employee accidents in certain specific work situations are general emotional maturity and the emotional conditions at the time of the accident. An analysis of the causes of accident susceptibility among 50 motormen of the Cleveland Railway Company, published by the Metropolitan Life Insurance Company, is reproduced in Fig. 18.6. In this analysis four single items—faulty attitude, impulsiveness, nervousness and fear, and worry and depression—account together for 32 per cent of the accidents among the group studied. Since these four items are essentially emotional in nature, it may be inferred that emotional conditions account for a considerable proportion of accidents. It is well accepted psychologically that many individuals vary in their general emotional state between a "high" and a "low" condition. These conditions are often

PRIMARY CAUSES OF ACCIDENT-PRONENESS

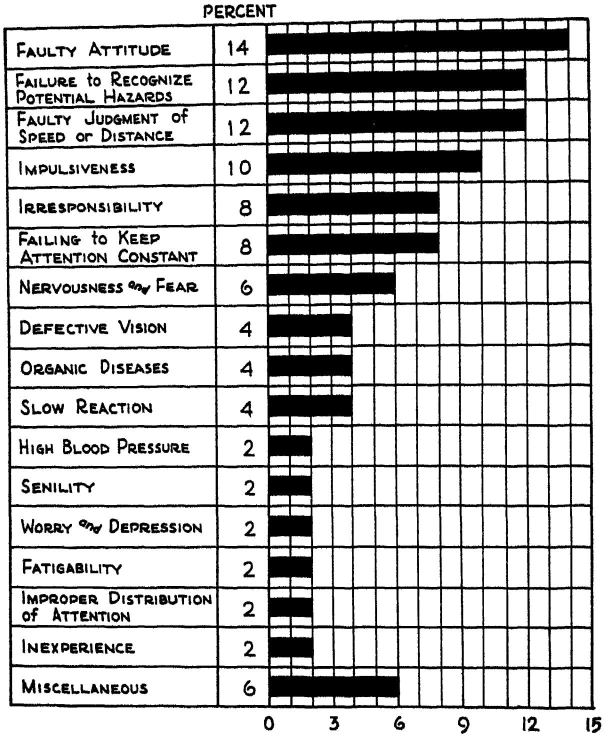


Fig. 18.6. Primary causes of accident proneness according to a survey among 50 motormen of the Cleveland Railway Company

fairly regular in their occurrence and appear in a cycle that ranges from one extreme to another. This situation is readily recognized in its extreme form, as it results in the so-called manic-depressive psychosis. In a minor form, however, it characterizes many people who are in no sense of the word psychotic, but who do, nevertheless, experience definite "ups and downs" in their general emotional state. The relationship between these emotional cycles and the frequency of accidents has been shown in an investigation by Hersey,[28] who found that the average worker is emotionally low about 20 per cent of the time and that more than half of the 400 minor accidents studied occurred during these low periods. According to

[28] R. B. Hersey, "Emotional Factors in Accidents," *Personnel Journal,* 15 (1936), 59–65.

chance alone only 20 per cent of the accidents would have occurred during such periods; the differential may therefore be taken as indicating the effect that an emotional depression has upon an employee's safety.

The general importance of the employee's emotional condition is further attested in another study by Hersey[29] in which it was reported that the production of industrial workers is around 8 per cent higher during periods when the men are elated, happy, hopeful, and cooperative than when they are suspicious, peevish, angry, disgusted, pessimistic, apprehensive, or worried. A favorable emotional condition thus not only is desirable from the standpoint of safety but also from the standpoint of plant productivity.

Supervisors and others dealing with human relations problems have long been aware that certain individuals vary from day to day in their agreeableness or ability to get along with others. Too often it has been the practice to handle such individuals either with a sharp reprimand or a general comment to the effect that so and so has a "grouch on" today. An employee who periodically gets a "grouch on" very likely is one who is afflicted with more than the average amount of cyclic fluctuation in his emotional state. To reprimand such an employee usually does about as much good as to reprimand a man with a broken arm for not doing his full share of work. Such an individual should be dealt with clinically to unearth the cause of his emotional condition. If this is impractical, he should be kept off a job that is in any way hazardous for the duration of his unsatisfactory emotional state. It should be emphasized that an employee who suffers from extreme variations in mood is not doing so for the fun of it. He cannot help himself, and he, himself, is unhappier as a result of his condition than are any of the people who are in contact with him.

Where there is any inkling of emotional problems with an employee, it would of course be desirable for him to place himself in the hands of a professionally trained person such as a psychiatrist or a clinical psychologist.

Mental ability. Mental ability is another personal factor that has been found in some circumstances to be related to accident frequency For some jobs a certain minimum mental ability level is required if an employee is to escape the hazards that occur on the job. In one situation, for example, Chambers[30] found that very few accident-prone individuals were above average in handwork, intelligence and learning ability, dependableness, and industry. The accidents for the most part happened to individuals who were relatively low in the traits associated with mental

---

[29] R B. Hersey, "Rates of Production and Emotional State," *Personnel Journal,* **10** (1932), 355–364.

[30] E. G. Chambers, "A Preliminary Inquiry into the Part Played by Character and Temperament in Accident Causation," *Journal of Mental Science,* **85** (1939), 115–118.

ability. In apparent contradiction to such implications, however, other studies have revealed no correlation between degree of mental ability and accident frequency. The disagreement is more apparent than real, however, for investigations of this latter type have attempted to correlate the degree of intelligence with accidents among employees who are above a certain minimum intelligence level. In most work situations it seems clear that, above a minimum critical mental ability level, little, if any, relationship exists between further amounts of mental ability and freedom from accidents. But in many work situations it still remains quite necessary for an employee to possess this minimum amount of mental ability. Those who lack it are quite likely to be hazardous employees. It is primarily for the identification of this extremely low group that the use of mental ability tests is recommended as an important part of any accident-prevention program.

Ratio between Perceptual and Muscular Speed An interesting hypothesis that seems to account for at least a part of the cause of accident proneness of employees in some kinds of work has been advanced by Drake.[31] Drake administered to a number of accident-prone and safe individuals a series of psychological tests. The tests were roughly divided into two groups, namely, those dealing with perception and those dealing with muscular responses. The perceptual tests were primarily concerned with visual discrimination and the muscular tests with the speed of executing a number of routine manual activities. In the course of examining these test results it was observed that the accident-prone persons tended to have motor test scores that were relatively higher than their scores on the perception tests. It was also observed that the safer employees or accident-free individuals tended to have motor test scores that were lower, relatively, than their scores on the perception tests.

By a procedure that need not be described here, a composite score was derived that, in effect, indicated, relatively, the difference between each person's perceptual and muscular speed scores. This composite score, in turn, was related to accident indexes of 38 people, as shown in Fig. 18.7. The general trend of the relationship is clear. Though it is not a straight line or linear relation, individuals with negative composite scores show a definite tendency to be among those with the high accident indexes, whereas those with the positive composite scores are relatively free from accidents. Drake's statement of the principle involved, as illustrated in Fig. 18.7, is that "Individuals whose level of muscular action is above their level of perception are prone to more frequent and more severe accidents than those individuals whose muscular actions are below their perceptual level. In other words, the person who reacts quicker than he

[31] C. A. Drake, "Accident Proneness: a Hypothesis," *Character and Personality,* 8 (1940), 335–341.

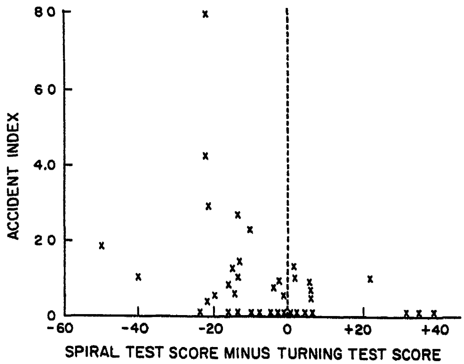


Fig. 18.7. Relation between accident proneness and test scores dealing with perceptual and muscular speed.

can perceive is more likely to have accidents than is the person who can perceive quicker than he can react."

Other Psychological Factors Other personal factors have also been found, in some situations, to be related to accident frequency. Among such factors are motor performance, and personality and temperament factors. In an investigation by Speroff and Kerr[32] 90 men on the finishing end of a steel mill hot strip were studied with reference to accident records and interpersonal preferences. It was found that workers who were most liked by fellow workers tended to be accident free while most of the high accident rates were found among the men most disliked by their fellow workers. Using a nomination method by fellow workers to measure the extent to which a man was liked by his peers, a correlation of —.54 was found between number of nominations and number of accidents during the period studied.

Identifying Factors Related to Accident Proneness Where accident proneness has been found to be characteristic of some of the employees on a job, there must be some personal variable or variables that are contributing to the consistently higher accident rates of some of the employees. While some such variables have been mentioned above, they are only suggestive; there undoubtedly are others, such as lack of training. It has not been the intention to imply, however, that the same assortment of personal variables would be the causative factors on all jobs. Rather, the point emphasized is that the personal variables that contribute to accident proneness on one job may be unique to that job. Therefore, in each work situation it is necessary to identify, by appropriate statistical

---

[32] B. Speroff and W. Kerr, "Steel Mill 'Hot Strip' Accidents and Interpersonal Desirability Values," *Journal of Clinical Psychology*, 8 (1952), 89–91.

analyses, what factor or factors are associated with accident frequency on the job in question.

HOW TO REDUCE ACCIDENT EXPERIENCE

Where it has been found that accident frequency is related primarily to the work situation, the obvious attack should be directed toward the particular features of the situation that seem to be associated with the accident frequency. Various schemes are used, such as by installing protective guards on machines, redesign of machines, change of methods, etc.[33]

Where accident proneness is a major factor in accident frequency, however, efforts to reduce accidents must be focused on the *employees* rather than on the *work situation.* The first step in this process is the *identification of the factors associated with accident proneness,* such as discussed above. Once such factors have been identified, it is posible to consider the development of plans directed toward the *specific* aspects of the problem that offer some promise of reducing accident frequency.

Placement of Employees When it has been found that accident-prone employees and those with good accident records differ significantly on some trait or characteristics, it is possible to place on the job in question those persons who have the same characteristics as those of the present employees with good accident records. Frequently this can be done with appropriate psychological tests such as vision tests, mental ability tests, or tests of other types.

Rigid adherence to a policy of such placement is as beneficial to the employees as to the management. When employees are placed without reference to those psychological and personal factors that are related to accident proneness, the end result is the placement of employees in such a manner that an unnecessary number of accidents occur. For example, it has been demonstrated (see page 567) that on some jobs employees who fail certain vision tests experience more industrial accidents than employees who pass those tests. A comprehensive visual performance test is necessary as a part of employment procedure if the employees who are accident prone for visual reasons are to be identified. Where other personal variables are found to be related to accident frequency, the use of appropriate tests or other predictors would of course be in order.

Safety Training In the case of new employees, it may be appropriate to set up training programs before they begin their jobs, in order to familiarize them with possible hazards, and with those practices and methods that will minimize the likelihood of their having accidents, and

[33] Various sources deal with general practices and specific techniques of making the working situation safer, such as Heinrich, *op. cit.*, and Blake, *op. cit.*

to develop proper attitudes toward safety practices. Training programs are also sometimes appropriate for present employees. In such training programs it is, of course, important to apply sound principles of learning, such as discussed in Chapter 10.

Walker and Potter[34] have emphasized the importance of worker participation in safety training. They point out that if employees are encouraged to participate in defining the safety requirements for jobs, the rules formulated will be accepted by them and provide a basis "for the constant retraining and reminding which the workers require" about safety.

An excellent example of the reduction in lost-time injuries which followed an intensive training program on safety teamwork has been reported by Ewell.[35] The program was started in 1930, a year in which the Procter and Gamble Company experienced 30 lost-time accidents per million man-hours worked. As shown in Fig. 18.8, the lost-time-accident

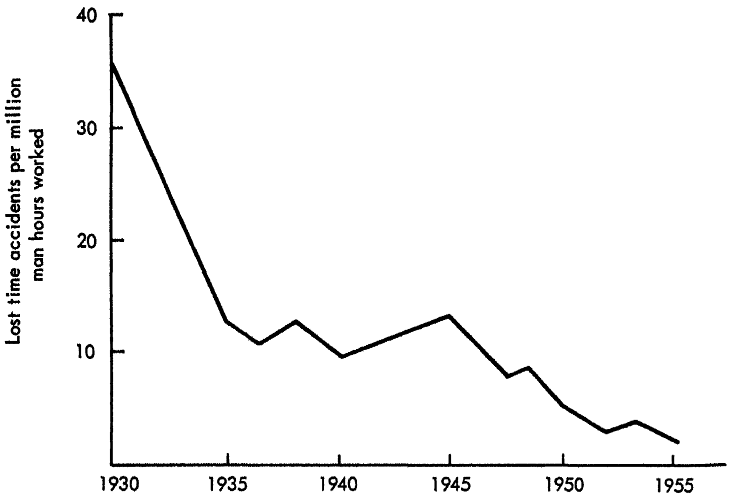


Fig. 18.8. Lost time injury reduction following a training program in 20 plants of the Procter and Gamble Company.

rate shows a consistent downward trend until, in 1955, only 1.01 lost-time accidents per million man-hours worked were experienced. The data also showed that in 1955, 11 of the 20 plants involved had perfect safety records.

---

[34] W. S. Walker and C. J. Potter, "Worker Participation in Safety Through Job Analysis," *Personnel,* 31 (1954), 141–147.

[35] J. M. Ewell, "1955, A Yardstick for '56," *Procter and Gamble Safety Bulletin,* January, 1956.

Use of Posters Posters and placards may help the new employee to some extent, *if* the posters tell him specifically *what* to do and what *not* to do, and if he reads the posters and remembers what he reads. But too often posters simply show an accident and, in large type, say "Be Careful." Since only the maladjusted individual is likely to hurt himself voluntarily, the degree of carefulness of the ordinary employee is determined by (1) what he knows of safe and unsafe practices, and (2) how important he thinks it is to follow his knowledge on the subject. In other words, some employees follow unsafe practices because they actually do not know any better. The need in such cases is for specific information on what to do and what not to do. A systematic and comprehensive series of conferences covering this specific information is a much more effective means of informing the employees than is the chance and casual learning obtained from posters or from observations that employees may (or may not) make. Once the information has been given to the employees, still further education is necessary to make sure that they *believe* as well as *know* that certain practices are unsafe.

One investigation on the effect of using safety posters has been published by Laner and Sell.[36] They report that when safety posters were used, more safe behavior was observed during a 6-week test period at 7 British steel plants. In 7 of 13 shops, safe behavior increased by more than 20 per cent.

Personal Protection For some kinds of work, various types of personal protection devices are used to minimize the possibility of injury. Such devices include safety goggles, safety shoes, safety belts and harnesses, special clothing, helmets, etc. While employees frequently resist the use of such devices, through training, posters, and other methods of communication it is frequently possible to reduce the resistance to their use.

---

[36] S. Laner and R. G Sell, "An Experiment on the Effect of Specially Designed Safety Posters," *Occupational Psychology*, 34 (1960), 153–169.

VI

Psychological Aspects of Consumer Behavior

All people are "consumers" of the goods and services of our economy. The behavior of people in their roles as consumers is a legitimate area of interest for those interested in the study of human behavior. In considering the area of consumer behavior, it is suggested that the consumer relationship be considered as any relationship between an organization that provides goods and services, and the individuals who are recipients thereof. While this usually is viewed in the context of consumer goods produced by private industry, it is equally reasonable to consider people as consumers of the services of many other types of organizations, such as educational institutions, governmental organiza-

tions, and the like. The operations of an economy are intended to serve the needs of people. It is suggested that consumer behavior be viewed in this broad framework, that of the utilization by people of the goods and services of our entire economy.

19

Consumer Psychology

Everyone is a consumer of at least some of the goods and resources of our economy. For our purposes, however, let us stretch the concept of "consumer" from its usual connotation relating to the consumption of the goods and services of private industry. Let us also think of people as consumers of the services of other types of organizations, such as educational institutions, government organizations (local, state, and federal), hospitals, religious organizations, and nonprofit organizations, as well as others. We, the public, are "users" of the services of the policeman, the mail carrier, the minister, the teacher, yes, even the tax collector, in a manner that is somewhat analogous to those services offered by the telephone company, the toothpaste manufacturer, the local retailer, the physician, or the loan company. The consumer relationship, then, can be considered as any relationship between an organization that provides goods or services, and the individuals who are the recipients thereof.

Consumer research is concerned with the systematic study of the facets of this relationship. While there are, of course, many aspects of consumer research, a few of the more important ones are the following: studies of consumer preferences (taste, style, features of products, etc.); product testing; consumer attitudes and motivation; buying habits and

patterns; brand preferences; media research (such as composition of audiences reached through TV, radio, newspapers, magazines, and other advertising media); effectiveness of advertisements and TV commercials; packaging; estimating demand for products or services; economic expectations of people; and studies of "images" of products. In this chapter we will be concerned with some of these aspects of consumer research.

A FRAME OF REFERENCE

In approaching the study of consumer behavior, the question of one's point of view or objectives might be raised. By and large, the study of consumer behavior in the past has been largely from the point of view of selling to the public certain goods and services that are proffered. In this connection, it is probable that these efforts have generally contributed to the development of the present-day economy and standard of living. In this process, there have, unfortunately, been at least some circumstances in which the practices of purveyors of goods and services have raised questions of ethics and good taste. The reader undoubtedly can call to mind examples of, say, advertising or sales practices that are out of bounds by usually acceptable standards of ethics or good taste, or that might be considered unwarranted "manipulation" of consumers. Such practices, in turn, might cause one to wonder about the possible ethical aspects of research relating to consumer behavior. In this connection we would like to point out that knowledge as such has no positive or negative valence in terms of moral, ethical, or good-taste considerations. It is the *application* of knowledge that takes on implications in terms of value systems. Knowledge about human behavior can be used to the advantage or disadvantage of mankind in many aspects of life, such as in politics, international relations, education and training, and personnel placement, as well as in the consumer domain.

With respect to the application of knowledge of consumer behavior, Perloff's [1] comments seem to offer a very appropriate frame of reference:

> Consumer benefits . . . would doubtlessly be multiplied if the psychologist should seek *directly and explicitly* to serve the consumer's needs, to study the consumer qua consumer, as it were, not as an individual whose attention and purchasing behavior are coveted to serve ends, the propriety and economic value of these ends notwithstanding, determined by advertising and the mass media. What I would like to propose, therefore, is that a relatively new frontier the industrial psychologist might scout would be that in which he seeks to study the consumer for the sake of understanding consumer behavior because consumer behavior is scientifi-

[1] R. Perloff, "Potential Contribution of the Consumer-Oriented Psychologist," *Business and Society*, 4 (No. 2, Spring, 1964), 28–34

cally important on the one hand, and is relevant to helping the consumer derive greater satisfaction and pleasure from the products he consumes, on the other.

THE BASIS OF CONSUMER BEHAVIOR

It has been pointed out by Katona[2] that there are five sets of variables that are relevant to consumer behavior. In the first place there are *enabling conditions* that set the limits to the consumer's discretion: his income, assets, and access to credit. Second, his economic behavior is influenced by *precipitating circumstances* such as increase or decrease in purchasing power, a change in family status, a move to a new house, etc. Third, *habit* plays an important role, especially in the purchase of such items as foods. Fourth, the *contractual obligations* of people affect their economic behavior, as, for example, rent, life insurance premiums, taxes, installment payments on automobiles, etc. The fifth factor is the consumer's *psychological state.*

The behavior of people in their roles as consumers is, of course, predicated upon essentially the same factors as is their behavior on their jobs or in other facets of their lives. Among the more important of these are motivation, perception, and learning. While these have been discussed in earlier chapters, brief mention will be made of them in our present context.

Human Motivation As indicated in Chapter 12, people have *needs* which they seek to fulfill. *Incentives* are those things which are perceived as fulfilling *needs.* Human behavior consists of those actions that occur in the process of attempting to achieve the incentives that are sought. In the consumer context, the goods and services that people buy, or avail themselves of, are the incentives which they perceive as being capable of fulfilling their needs. In the earlier discussion of motivation, Maslow's[3] hierarchy of needs was presented. To the extent that this hierarchy has validity, it is reasonable to postulate that consumer behavior is dominated to a considerable extent by it, in the sense that the kinds of goods and services that people seek would be related to the needs that are implied by this hierarchy. The starving individual would, of course, be primarily concerned with obtaining food. In turn—and depending upon the general level of the needs that have been fulfilled—individuals would seek those goods and services which they perceive as being capable of fulfilling their next-higher order of need, such as for personal safety.

While all behavior can be viewed as being the result of motivation,

---

[2] G.Katona, "Economic Psychology," *Scientific American* (October, 1954).

[3] A. H. Maslow, *Motivation and Personality* (New York: Harper & Row, Publishers, 1954).

any single act typically involves several or many different motives, frequently operating in opposite directions. Because of the recognized importance of motivation in consumer behavior, some years ago there was considerable interest in what is called motivation research. Such research is directed toward deriving information about the motivational bases of consumer behavior.

Methods of study of motivation. As pointed out by Ferber and Wales,[4] the study of human motivation in the consumer context can be carried out in two somewhat different ways. One of these is essentially a clinical, individual approach, involving the use of depth interviews, projective devices, personality tests, etc., in order to learn something about individuals, and the motivational forces that presumably dominate their behavior. On the basis of such clinically obtained data and inferences, from samples of individuals, one might make some admittedly subjective assessments about the motivation of people as related to some product or service.

The other approach is essentially a statistical one, based on responses to survey questionnaires or fairly well-structured interviews, or based on some type of behavior that can be observed. While the statistical data consist of responses to the questions that are used, or observations of the behavior, inferences frequently can be made from such data about motivational variables. For example, in a study carried out by the Treasury Department during the war relating to the purchase of War Bonds, it was found that personal appeals to individuals were more effective than public appeals by radio, newspaper, etc. But when bond buyers were asked their reasons for buying, very few replied that it was because of the personal solicitation. Rather, they tended to give such reasons as helping the soldiers, personal savings, etc.

In discussing motivation research, Dichter[5] argues strongly for the first of these approaches, in order to identify the major psychological variables and personality elements which may have a dynamic effect on consumers' attitudes. He suggests, for example, that such factors as adventurousness, insecurity, prejudice, and self-indulgence often are of more significance in determining market behavior of people than such objective factors as income, age, etc. On the other hand, it has been pointed out by Cannell[6] that a statistical approach provides a more accurate picture of the public than do the more insightful but limited techniques of clinical studies. He emphasizes the point, however, that techniques are measurement devices only, and are not theory; theory

[4] R. Ferber and H. G. Wales, *Motivation and Market Behavior* (Homewood, Ill.: Richard D. Irwin, Inc., 1958).

[5] Ernest Dichter, "Toward an Understanding of Human Behavior," in Ferber and Wales, *op. cit.*

[6] C. Cannell, "A Psychologist's View," in Ferber and Wales, *op. cit.*

is necessary to understanding the consumer. With respect to methods as such, it is probable that both of the approaches discussed above have their place in consumer research.

Perception How people view themselves and others, and the goods and services that are available to them, have significant effects on their behavior as consumers. Mention was made above, for example, of the *incentives* as being those things which are perceived as fulfilling *needs.* Whether—to an outsider—a given incentive would seem to be relevant to fulfill the "need" of another individual is beside the point. The individual's behavior will be influenced by his *perception* of whether something will, or will not, fulfill his need. Thus, if an individual perceives a sports car as fulfilling his need for status, he will behave accordingly (perhaps buying such a car), even though, in the eyes of others, his status is not so changed. In fact, others might perceive such an act as, say, an indication of "show-off" behavior. Many "perceptions" of people, then, stem from their own unique motivations and frames of reference.

Aside from such individualistic aspects of perception, however, there are certain patterns of perceptual processes that are more generalized. Some of these have considerable importance to various phases of consumer behavior. In advertising, for example, certain features of advertisements are more consistently perceived than others. Or in product design, certain features may generally be perceived as more adequately fulfilling certain "needs" than others.

Learning Learning is another important factor that influences the behavior of people in their roles as consumers. Habits, for example, are manifestations of learning. Thus, to the extent that people buy products, or avail themselves of services, on an habitual basis, their behavior is "learned"—for better or for worse. Another example of the implications of learning on consumer behavior is that of remembering the experiences they have had with various products or services. If, for example, a farmer has had difficulty starting a particular tractor in the winter, this "learning" may influence his future tractor selection. Another example of the implications of learning in the consumer context is the extent to which people remember (recall) things to which they have been exposed, by TV, newspapers, magazines, or other media.

With respect to learning in the field of consumer psychology, Lucas and Britt [7] call particular attention to the distinction between *recognition* and *recall.* Many psychological investigations have demonstrated that people tend to forget rapidly, and that the speed of forgetting accelerates. In part because of this, the measure of learning (at various

[7] D. B. Lucas and S. H. Britt, *Measuring Advertising Effectiveness* (New York: McGraw-Hill Book Company, 1963).

points in time) depends on whether one measures recognition or recall. In consumer research, recognition typically is measured by presenting subjects with advertisements, magazine articles, or other relevant material, and asking them if they can identify the material as having been seen (or "noted") previously. Sometimes a distinction is made between *noting* (seeing) and *read most* (thorough reading).

Recall, on the other hand, usually is measured by asking people to reconstruct their impressions, or to give information about, articles, advertisements, or whatever other material is being tested. Aside from the cues that are used to direct the attention of the respondent to the particular material in question, the method places a heavy burden upon memory of the content of the material. This usually is facilitated by trained interviewers who probe with such questions as: "Can you tell me about the article?" "What did the advertisement look like?" and "Can you be more specific?" The verbatim responses are recorded and subsequently analyzed.

Comparison of recognition and recall of advertisements. A study of retention (or loss of memory) of information about advertisements is reported by Lucas.[8] In this study he distinguished between *recognition* (in terms of "noting" scores) and *"aided recall."* In the study it was found that recall tended, relatively, to decline over a period of days, whereas recognition was maintained for at least a couple of weeks. This comparison is shown in Fig. 19.1 for a period of seven days.

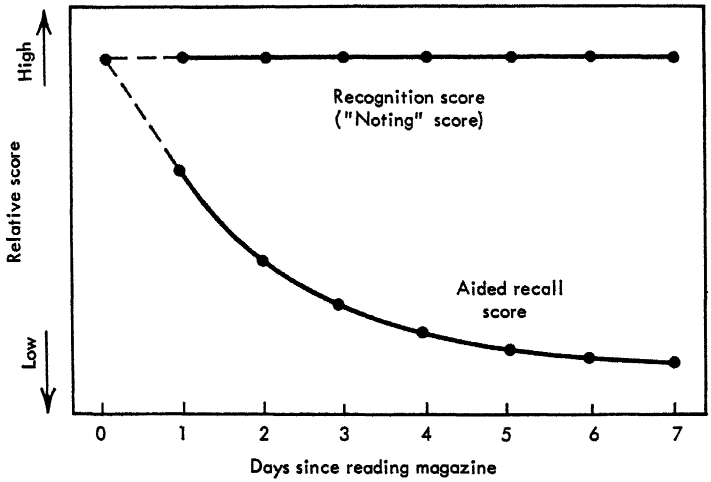


Fig. 19.1. Changes in recognition ("noting" scores) and aided recall of 96 advertisements in a national magazine (600 subjects). (From Lucas, *op. cit.*, Fig. 1.)

[8] D. B. Lucas, "The ABC's of ARF's PARM," *Journal of Marketing*, 25 (No. 1, July, 1960), 9–20.

Recall and recognition of magazine articles. As an illustration of a consumer study relating to learning, Greenberg and Garfinkle[9] report an investigation of the recall of information relating to 60 articles in 4 magazines, namely, *Life, Look, Saturday Evening Post,* and *McCall's.* We will report, in particular, the results of an analysis of recall with a sample of single women, ages 17 to 24, as related to the amount of space allocated in the articles to "visual" material. For each article, the percentage of the total space used that was devoted to visual material was determined, the visual material including pictures, captions, etc. (as opposed to printed text material). The 60 articles were then divided into 3 groups on the basis of these percentages, namely, high visual content, medium, and low.

Each respondent was questioned about a representative sample of the total sample of 60 articles (the total sample including 6 articles in each of 3 magazines and 12 articles in the other). Readership and retention were measured as follows:

1. Unaided recall. The interviewer read the title of the article and asked the respondent if she remembered reading it, and if so to describe what she remembered. Correct identifications were classified as "unaided recall."
2. Recognition. Next the interviewer opened the magazine issue to the article in turn, and all issue readers were asked which ones they remembered reading. All such claims were classified as "recognition."
3. Correct answers on content. Three questions were asked about each article. If the respondent answered two correctly (by correct "yes" or "no" responses), she was classified as giving correct answers on content.

Some of the results are given in Table 19.1. It can be seen that recognition was related to the amount of visual content, but, more relevant to our present interest, it can be seen that—of those who claimed to have read the articles—the percentage of "unaided recall" and the percentage giving "correct answers" were higher in the case of the articles with greater visual content, thus indicating that "learning" was greater when the articles were more profusely illustrated, or had higher percentages of other visual material.

Recall of advertisements as related to interests. With respect to recall of advertisements, Buchanan[10] reports a carefully carried out study in which the interests of respondents in eight products was measured by a mail questionnaire. Ten days later the subjects were "exposed" to advertisements in the consumer laboratory, in the case of four products to ads in a "dummy" magazine, and in the case of the other four to TV commercials. By an unexpected telephone call the

[9] A. Greenberg and N. Garfinkle, "Visual Material and Recall of Magazine Articles," *Journal of Advertising Research,* 3 (No. 2, June, 1963), 30–34.

[10] D. I. Buchanan, "How Interest in the Product Affects Recall: Print Ads vs. Commercials," *Journal of Advertising Research,* 4 (No. 1, March, 1964), 9–14.

Table 19.1

RECALL, RETENTION, AND LEARNING OF MAGAZINE ARTICLES VARYING IN AMOUNT OF VISUAL CONTENT, ON PART OF FEMALE RESPONDENTS *

	Visual content		
	High	*Average*	*Low*
Number of articles	29	20	11
Recognition (claimed readership)	70%	55%	43%
Unaided recall (as per cent of claimed readership)	54%	47%	40%
Correct answers (learning) (as per cent of claimed readership)	60%	53%	42%

* *Greenberg and Garfinkle,* op. cit, *from Tables 3 and 5*

next day the respondents were asked to recall these advertisements. The recall of the magazine advertisements was found to be significantly related to the previously measured "interests" in the various products. This was not the case with the TV commercials, although there was no indication as to why the difference existed between the two media—newspapers and TV.

Recall as related to behavior or attitudes. In connection with measures of recall, Haskins [11] surveyed 17 published studies which bear on the correlation between factual recall of information from advertisements on the one hand with attitude or behavioral changes on the other—the contexts of these studies varying considerably. Of these, two showed a relationship between recall of information attitude change in the expected direction, two in the unexpected direction (these boomeranged), and 13 showed virtually no relationship between what a person learned, knew, or recalled, on the one hand, and what he did or how he felt on the other. While the implications of this review are somewhat ambiguous, there seems to be no basis for inferring that the amount of recall of advertisements is necessarily related to consumer behavior or attitude changes.

Consumer Decisions Factors that enter into the decisions that people make are of course central to an understanding of consumer behavior. Motivation, perception, and learning, as discussed above, are significant ingredients of such decisions. In taking something of an overview of consumer decision-making, Katona [12] points out that a major finding of the consumer studies of the Survey Research Center

[11] J. B Haskins, "Factual Recall as a Measure of Advertising Effectiveness," *Journal of Advertising Research,* 4 (No. 1, March, 1964), 2–8.

[12] G. Katona, "Consumer Behavior Surveys and Marketing: A Point of View," in *Psychological Research on Consumer Behavior* (Ann Arbor, Mich.: The Foundation for Research in Human Behavior, 1962.)

may be summarized by simply saying that the consumer is sensible; while impulsive behavior does exist, it is much less frequent than is sometimes postulated.

In discussing the decision processes he differentiates between those consumer situations in which consumers make genuine decisions as opposed to acting in a quasi-automatic, habitual manner. He points out that the consumer neither can, nor should, consider every action he takes as a problem which requires deliberation, weighing of alternatives, and solution. Rather, we typically act as we have before under similar circumstances; this goes back to the matter of learning discussed earlier. On the other hand, there are circumstances under which genuine decision-making takes place, with consideration and weighing of various alternatives. Such decision-making probably tends to occur in the case of expenditures which are subjectively thought to be major and which are fairly rare, where there have been unsatisfactory past experiences, sometimes in the first purchase of a product, etc. As Katona [18] points out, when genuine decision-making takes place, consumers are not marionettes that can be manipulated. The efforts of the Federal Government and certain consumer organizations in the consumer education area would of course have the effect of furthering genuine decision-making on the part of consumers.

DATA COLLECTION METHODS

In the various types of consumer research, data are obtained regarding many aspects of consumers and their behavior and responses. One of the major aspects is the collection of data relating to consumer attitudes, opinions, motivation, values, preferences, buying habits, and other variables. Information about consumers and their opinions, motivations, values, preferences, etc. can lead to the development or modification of products or services that would meet more adequately the needs of consumers, and to the development of advertising programs that are more realistically related to such needs.

Some of the research methods used in these phases of consumer research are essentially the same as those used in other contexts that have been discussed earlier. These include interviewing procedures (discussed in Chapter 4), tests and measurements (Chapter 5), attitude and opinion measurement (Chapter 11), and various types of rating procedures (such as discussed previously in the context of performance appraisal, Chapter 9). Some of these methods have rather special or unique twists in the consumer field. However, a few of the more common methods will be mentioned, and in some cases illustrated.

---

[18] *Ibid.*

Requirements for Adequate Data Collection Before discussing methods as such, it is important to mention two critical points that will influence the adequacy of the resulting data and the extent to which results can be extrapolated to populations generally.

Adequacy of method. One of these requirements relates to the adequacy of the method used, such as a questionnaire, an interview procedure, etc. No fancy cover on a report is going to make up for inadequacies of the questionnaire, the interviewing procedures, or the lack of ability and training of interviewers, or inadequacy of experimental or statistical methods.

Sampling. If one is to collect information from a sample of people and would like to extrapolate from that sample to a population, the nature of the sample can be a major factor in the validity of any such extrapolation. In consumer research, the investigator typically has some "population" in mind. While this might comprise the population in its entirety, it usually consists of some restricted segment, such as housewives, car owners, adult males, families with children, retired individuals, voters, television viewers, etc.

Given the population—however it may be defined—there are basically two methods of sampling that can be used. One of these is *random* sampling; the other is *stratified* sampling. In random sampling, the sample should be so selected that each individual within the population has an equal chance of being selected for the sample. In stratified sampling, the population is first categorized in terms of certain presumably relevant characteristics, such as age, sex, marital status, education, income, etc., in order to know what proportion of the population falls within each category. The sample includes proportionate representation from the various categories. For example, if 25 per cent of the population are home owners, 25 per cent of the sample should be home owners (if home ownership is one of the stratification variables). Within a given stratum, individuals in the sample are selected at random from those in the population who fall in that stratum.

It should be noted that for many surveys no pretense is made that the sample is, say, a random sample of the true population. For example, if one surveys a sample of car owners in Santa Fe, New Mexico (even though it is a random sample of car owners in that city), there is no assurance that their responses will be representative of the responses of car owners in Dubuque, Iowa; Boston, Massachusetts; Punxsatawney, Pennsylvania, or the country as a whole. To extrapolate from such a survey to, say, the car owners across the country, is, then, to some degree risky. In practical situations, however, it is sometimes necessary to take some such risks. One should, however, recognize whatever such risks there may be in such extrapolation, and be willing to accept the possible consequences of being wrong. Sometimes, however,

one can get some inkling of the risks. If in previous surveys, for example, random samples of car owners in the above (or other) communities generally responded in the same way, one would have more comfort in extrapolating to car owners of the country as a whole than if the responses in such communities differed materially.

As an example of the possible influence of sampling on the results of a survey, Eastlack[14] describes a telephone survey carried out with two sampling procedures. By one method (A) the telephone interviewer is given assigned telephone directory columns, and is instructed to "work up" each column, skipping busy or no-answer telephones until completing an interview in that column. In the other method (B) the telephone interviewer was given a pre-listing of names, with instructions to make four attempts to interview each respondent before giving up. Of the respondents interviewed by the two methods in one city, it was found that there was a difference in their "product usage" (of those interviewed by method A, 56 per cent reported the use of instant coffee, where as by method B it was 71 per cent). These differences strongly suggest that those who were not at home when phoned once (method A) tended to have some difference in their usage of instant coffee. While there is no specific indication from the data as to why this is so, it is possible that more of those who were not at home when phoned once might have jobs, and, in turn, tend to use instant coffee, than in the case of those who were at home. In any event, is is probable that method B resulted in a more representative sample than did method A.

Consumer Panels In the collection of data in consumer research, sometimes consumer panels or juries are used. Depending upon the purpose in mind, the members of the panel may be brought together as a group, dealt with individually, contacted by phone, contacted by mail, or used in some other manner. Such panels are used for various purposes, such as testing advertising copy, getting opinions about various products and packaging, and taste testing of food products. Certain of the television ratings are based on the use of samples of television viewers which might be thought of as a panel. In the case of one rating system, a device attached to the television set automatically records the programs to which the set is tuned in. The way in which the panel is used will, of course, depend on the purpose of the survey. The panel, in effect, is then a sample of people. Frequently the same panel is used over a period of time.

Questionnaire Surveys One of the most common methods used in consumer research is the consumer questionnaire survey method. This

---

[14] J. O. Eastlack, Jr., "Recall of Advertising by Two Telephone Samples," *Journal of Advertising Research,* 4 (No. 1, March, 1964), 25–29.

method consists of having questionnaires completed by a sample of people. The questionnaires may be distributed by mail, by hand, or by some other method.

Questionnaire development. As pointed out by Anastasi,[15] the responses to questions can be markedly influenced by the nature of the questionnaire and the questions, such as the grammatical form of the questions, the number of possible responses provided, their order, and their other characteristics. For example, Weitz[16] obtained significant differences in preferences in a survey relating to cooking ranges when the alternatives were presented verbally *versus* pictorially. In addition, emotionally loaded words in questions usually will influence the responses of subjects. Anyone constructing a questionnaire should be familiar with some of these pitfalls, and should follow sound questionnaire-construction practices.[17]

One important feature in questionnaire development is the form of the question used. This can have an influence on the responses obtained, and can therefore influence the interpretation of the results obtained. The influence of the form of a question on responses was illustrated by a comparison made by Belson and Duncan[18] of the responses of over 1,000 people to questions about the newspapers and magazines they had looked at yesterday and the TV programs they had watched yesterday. One-half of the subjects were given a checklist questionnaire, listing most newspapers and magazines and TV programs, with instruction to check off those that they had read or watched; there was also a checklist category for "all others" in each group. The other half received open-end questions and were asked to list the relevant items on the basis of their recall. The average checklist respondent claimed viewing half again as many programs as those using the open-end form. But on the other hand, the checklist form had a depressing effect on some free-recall items, especially for TV programs with incidental viewing or hard-to-remember names; newscasts, for example, were mentioned by only one per cent of

---

[15] Anne Anastasi, *Fields of Applied Psychology* (New York: McGraw-Hill Book Company, 1964), p. 271.

[16] J. Weitz, "Verbal and Pictorial Questionnaires in Market Research," *Journal of Applied Psychology,* 34 (1950), 363–366.

[17] The reader is referred to such sources as.

A. B. Blankenship (chairman of special committee of American Management Association), "Questionnaire Preparation and Interview Techniques," *Journal of Marketing,* 14 (October, 1949), 393–433.

H. W. Boyd, Jr., and R. Westfall, *Marketing Research: Text and Cases* (Homewood, Ill.: Richard D. Irwin, Inc., 1956).

M. Parten, *Surveys, Polls, and Samples: Practical Procedures* (New York: Harper & Row, Publishers, 1950).

S. Payne, *The Art of Asking Questions* (Princeton, N. J.: Princeton University Press, 1951).

[18] W. Belson and J. A. Duncan, "A Comparison of the Check-List and Open Response Question Systems," *Applied Statistics,* 2 (No. 2, June, 1962), 120–132.

the checklist respondents *versus* 25 per cent of the open-end respondents!

In the development of questionnaires, it is very strongly recommended that they be pretested on a small sample prior to their use on a wider basis. The pretest should be carried out with the same kinds of subjects as those who will complete the questionnaire later.

Biases in responses. An additional possible source of distortion in questionnaire responses is the subjects themselves. These biases may be generalized tendencies, or they may be associated more with certain categories of subjects. In a study by the Federal Reserve Board as reported by Rikumo,[19] for example, about 1,800 people reported their monthly payments on loans, and the amount of their loans. These reports were later compared with actual indebtedness. It was found that there was a significant tendency for people to report that they owed *less* money than they really did owe, by an average of about $83. Nearly half of the respondents reported lower values. There was, however, reasonable accuracy in reporting the amount of monthly payment. The underreporting was found to be particularly associated with the amount of debt (with underreporting being greater for those with higher indebtedness), and with education and occupation.

Another example of bias arising from respondents is reported by Wells.[20] In a previous study by Couch and Keniston [21] it had been found that certain individuals tended, when responding to personality and attitude items, to say "yes" and "agree," whereas others tended to say "no" and "disagree." Further, the "yeasayers" tended to be extroverted, voluble, impulsive, and excitable, whereas the "naysayers" tended to be introverted, careful, cautious, and controlled. In his study Wells [22] used an adaptation of the scale (the YN-2 Scale) that had been developed by Couch and Keniston [23] for differentiating between the "yeasayers" and the "naysayers," using this scale in connection with a consumer survey of 100 housewives regarding claimed possession versus (later-proved) actual possession of five national magazines. He found that of the "yeasayers" (as identified by the YN-2 Scale) 69 per cent made at least one unsubstantiated claim of magazine possession, as contrasted with 18 per cent of the "naysayers."

Rating and Scaling Methods. Closely related to questionnaires are various rating and scaling methods. Some such methods were discussed

[19] I. Rikumo, "An Analysis of Response Errors: A Case Study," *Journal of Business*, **36** (No. 4, October, 1963), 440–447.

[20] W. D. Wells, "How Chronic Overclaimers Distort Survey Findings," *Journal of Advertising Research*, 3 (No 2, June, 1963), 8–18.

[21] A. Couch and K. Keniston, "Yeasayers and Naysayers: Agreeing Response Set as a Personality Variable," *Journal of Abnormal and Social Psychology*, **60** (1960), 151–174.

[22] Wells, *op cit.*

[23] Couch and Keniston, *op. cit.*

and illustrated in the earlier context of performance appraisal (Chapter 9). Essentially the same techniques are also applicable in the context of consumer research, with, of course, different content. For example, it would be possible to have respondents rank, in order of importance to them, the factors they consider in, say, the selection of a house, such as type of neighborhood, convenience to schools, distance from public transportation, cost, style of architecture, etc.

Semantic differential. An additional type of scaling procedure that is used in consumer research (and other attitude research) is the semantic differential, described by Osgood, Suci, and Tannenbaum.[24] This technique deals with the determination of graduations in the semantic meaning of words when they are used to describe something. While it was developed for quite different uses than those in consumer research, it has potential utility in consumer research for such purposes as determining the "image" of products, institutions, media, etc. The method will be illustrated by an example. The example deals with a comparison of the images of radio, television, and newspaper media in Richmond, Virginia.[25] All available adults and teen-agers in a random sample of 447 households (drawn from newspaper route lists) were interviewed, using the scale that is illustrated in part in Fig. 19.2. Two months later, one-half as many households next door to those of the original sample were interviewed in a similar way concerning radio and television. A comparison of the results is shown in Fig. 19.3, this showing the values of a rating index that is based on the responses. The results generally indicate that the public views television as having more warmth than newspapers, but of newspapers being more intelligent, credible, moral, courageous, and reliable than television.

The semantic differential technique is one that can be adapted to various situations by selecting those pairs of words with opposite meaning that are considered to be relevant to the situation. For example, it was used as an index of attitudes toward two brands of food products, as reported by Barclay.[26] Without describing the details of the study, he found that the "brand attitude" scores of 199 women toward the two food products, as measured by the semantic differential method, differed significantly between those women who used one of the two brands (A or B) as contrasted with those who used the other or neither. While the differences were significant, however, they were small, suggesting that the

---

[24] C. E. Osgood, G. J. Suci, and P. H. Tannenbaum, *The Measurement of Meaning* (Urbana, Ill.: University of Chicago Press, 1957).

[25] *The Climate of Persuasion: A Study of the Public Image of Advertising Media* (Richmond, Va.: The Richmond Newspapers, Inc., 1959).

[26] W. D. Barclay, "The Semantic Differential as an Index of Brand Attitude," *Journal of Advertising Research*, 4 (No. 1, March, 1964), 30–33.

Your frank appraisal of — The Richmond News Leader ( ) / Richmond Times-Dispatch ( )

Think of this newspaper as though it were a person, a regular visitor to your home. Listed below are words to describe your feelings about the newspaper. By placing an X anywhere along each line, you indicate your feelings in each case. We want your honest, sincere opinion - so please be frank.

	0	1	2	3	4	5	6	7	8	9	10	
Aloof	O	O	O	O	O	O	O	O	O	O	O	Friendly
Untruthful	O	O	O	O	O	O	O	O	O	O	O	Truthful
Prejudiced	O	O	O	O	O	O	O	O	O	O	O	Unprejudiced
Immoral	O	O	O	O	O	O	O	O	O	O	O	Moral
Unintelligent	O	O	O	O	O	O	O	O	O	O	O	Intelligent
Dreary	O	O	O	O	O	O	O	O	O	O	O	Cheerful
Fearful	O	O	O	O	O	O	O	O	O	O	O	Courageous
Distant	O	O	O	O	O	O	O	O	O	O	O	Neighborly
Deceitful	O	O	O	O	O	O	O	O	O	O	O	Trustworthy
Superficial	O	O	O	O	O	O	O	O	O	O	O	Thorough
Intolerant	O	O	O	O	O	O	O	O	O	O	O	Tolerant

Fig. 19.2. Illustration of part of a semantic differential scale. This particular one was used in measuring the public image of newspapers; comparable scales were used for radio and television media. Only some of the pairs of words used are shown. (Courtesy of the Richmond Newspapers, Inc.)

method as used in this particular situation was not particularly sensitive.

Interview The interview is another method of obtaining information from, or about, people.

Structured interview. In a structured interview (such as might be carried out in a door-to-door marketing research survey) the interviewer follows a general predetermined interview procedure. This form of interview is somewhat akin to a questionnaire, except that the questions are asked orally. The questions used in such a situation should be framed with the same kinds of considerations in mind as in the case of questionnaires, and like them should be pretested.

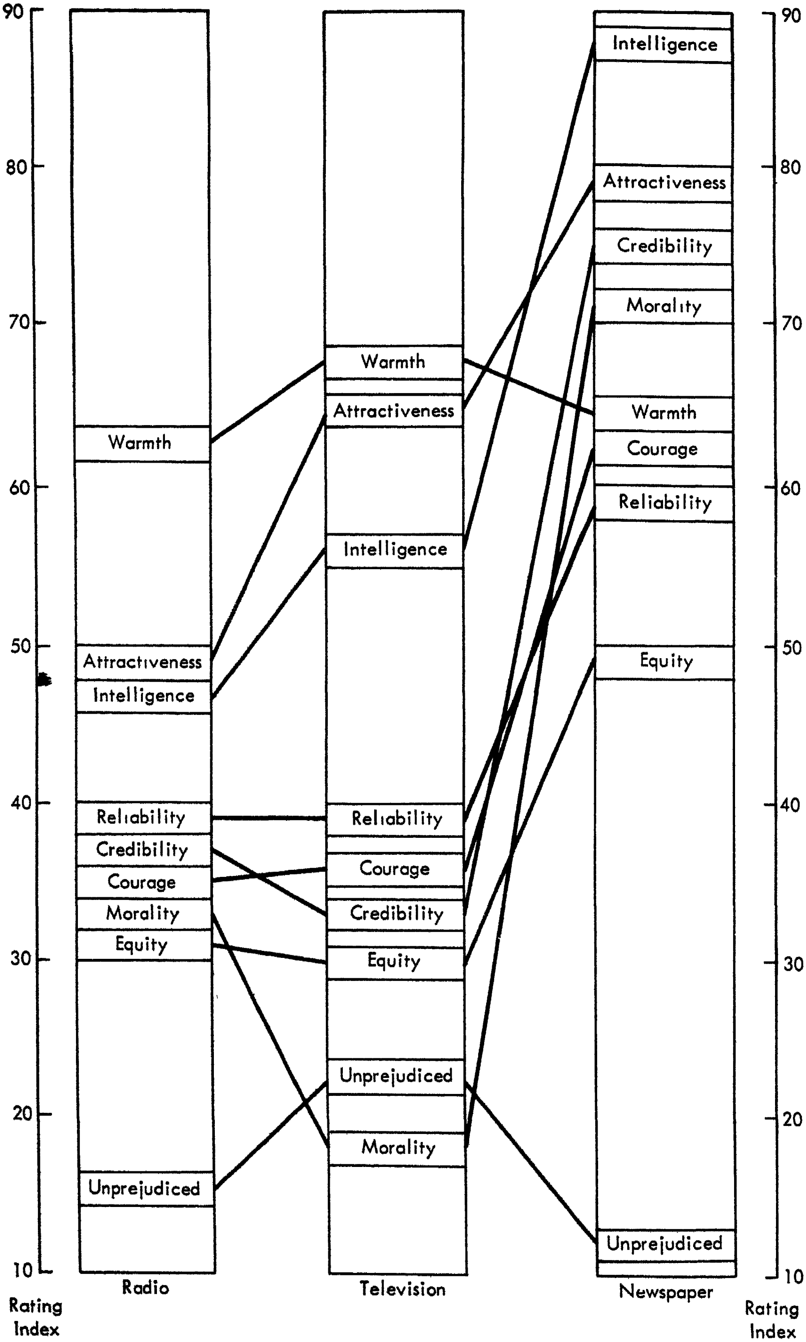


Fig. 19.3. Comparison of public images of three media as measured by the semantic differential technique The position of each category indicates its average rating index. (Courtesy of the Richmond Newspapers, Inc)

Unstructured interview. A more flexible interview procedure has been proposed by Gallup [27] for the purpose of obtaining information from customers which will reflect something about the influence of specific advertisements on their purchases. Since many people tend to view their purchasing decisions as their own (uninfluenced by advertisements), it has been found to be fruitless to ask people if such-and-such an advertisement influenced their purchasing decision. Gallup's interview approach starts *from* the purchase, and then proceeds *to* the advertising. The interviewee is encouraged to recall the circumstances leading to a given purchase and, in particular, to give details about any advertisements that may have had some part in the choice.

Still another form of interview is what Lucas and Britt [28] refer to as an informal, unstructured interview, such as in testing advertisements. For example, the testing of a comic-strip character for possible use in a hairdressing advertisement was based almost entirely on impressions given by interviewees while they were looking at the display.

Depth interview. Still another form of interview is the depth interview, which is used occasionally for gaining insight into human motivation as related to the use of products. As indicated by Dichter,[29] the depth interview is used in order to elicit the freest possible associations on the part of the respondent. The role of the interviewer is that of establishing rapport with the respondent, of starting the interview in a very general way, and of encouraging the respondent to express himself. The respondent generally determines the direction and content that the interview takes, thus giving the respondent the opportunity to talk about himself. Dichter [30] proposes the use of depth interviewing (and also of projective techniques, which are discussed later) under the following circumstances: (1) whenever the data being sought may not be present at the rational or conscious level; (2) whenever we are dealing with psychological mechanisms and not with simple cause-and-effect relationships; and (3) whenever the respondent has a chance to produce interference consciously or unconsciously between the time he understands the question and the time he answers it.

Advantages and disadvantages of interviews. One of the advantages of the interview method is that it usually results in obtaining data from a higher percentage of those in the sample. Dudycha,[31] for example,

---

[27] G Gallup, *Activation: A Major Development in Advertising Research* (Princeton, N. J.: Gallup and Robinson, Inc.), June, 1957.

[28] D. B. Lucas and S. H. Britt, *Measuring Advertising Effectiveness* (New York: McGraw-Hill Book Company, 1963), pp. 140–142.

[29] Ernest Dichter, "Toward an Understanding of Human Behavior," in Ferber and Wales, *Motivation and Market Behavior* (Homewood, Ill.: Richard D. Irwin, Inc., 1958).

[30] *Ibid.*

[31] G. J. Dudycha, *Applied Psychology* (New York: The Ronald Press Company, 1963), p. 362.

indicates that in interviewing surveys, interviews usually are completed with 80 per cent of the sample. With special efforts (including call-backs), it may be possible to contact successfully 90 to 95 per cent of the sample. With mail questionnaire surveys the percentage of returns usually is much smaller. In addition, the interview may offer such advantages as being able to establish better rapport with subjects, thus maintaining subject interest for a longer period of time (as when soliciting answers to many questions). This rapport, and the opportunity to clarify questions, can tend to elicit more accurate information. In the case of unstructured and depth interviews, it is possible to obtain data or impressions that could not be elicited by the use of more structured interviews or with questionnaires.

On the other hand, interviews are very time consuming and costly, which usually limits the number of individuals who can be surveyed—as contrasted to a questionnaire survey. In addition, there is the problem of selecting and training interviewers, and the very real risk of interviewer biases' influencing the resulting data in some unknown manner and degree.

Projective Techniques. Projective techniques—whether used in personality measurement, in consumer research, or otherwise—are methods of eliciting responses from people to intentionally ambiguous, unstructured stimuli. Since the stimuli are vague, it is presumed that responses made by subjects are "projections" of themselves. Interpretation of the responses by a trained individual might, therefore, reveal information indirectly about the subject that it might not be possible to obtain directly. Projective techniques have been used primarily in the field of clinical psychology. They have been used in consumer psychology primarily to derive inferences about motivation, attitudes, and other variables that are relevant to consumer behavior. A few of the specific techniques that have been used in consumer psychology will be mentioned briefly.

Association techniques. In these procedures, the subject responds to some stimulus with some word or expression that the stimulus brings to mind—that is "associated" with the stimulus. An example of this procedure is a word-association test reported by Lucas and Britt.[32] This test was used to study people's associations with certain key words of advertisement copy that related to the promotion of butter. The five key copy words (buttery, snack, sizzling, bun, and aroma) were fitted into a larger group of words, as follows:

1. table ________	5. buttery ________	9. red ________
2. orange ________	6. cold ________	10. bun ________
3. snack ________	7. aroma ________	11. friend ________
4. white ________	8. owl ________	12. sizzling ________

[32] Lucas and Britt, *op. cit.*, p. 143.

Every word in the list, and especially the sequence of words, has an influence on each response. This cannot be avoided, although the sequence can be varied or other background words substituted. The above pattern, used with hundreds of convenient subjects, produced such responses to the test as follows:

snack: tidbit, brunch, quick, hungry
buttery: smeary, fatty, greasy, yellow, heavy
aroma: smell, pleasant, odor, perfume
bun: roll, bread, gravy, oven
sizzling: hot, sputtering, spattering, meat

In the case of this particular example it was found that there were very few "taste" associations with *buttery,* possibly because of lack of relevant words in the English language. The results of association tests in advertisement copy writing can serve to guide the selection, and avoidance, of words for use in advertisement copy. Association techniques can also be used in other aspects of consumer research, such as the more clinically-oriented setting of depth interviewing, or in the measurement of personality variables. In this latter connection, for example, the Rorschach ink blot test is an example of an association type of test, since the subject is asked to indicate what he "sees" in (associates with) each ink blot.

Sentence completion tests. Sentence completion tests consist of parts of sentences, with certain words or phrases omitted. The omissions are of such a nature that the words or phrases filled in by the respondents have a significant influence on the meaning of the sentence, thereby making it possible to derive some inferences about the respondents.

An example of a sentence-completion test is reported by MacLeod[33] in a study of the expressed choices relating to fictitious makes of automobiles. The test consisted of 40 statements such as the following examples:

The best kind of car is one that ______________________.
The most enjoyable thing when you go to buy a car is __________.
Driving very fast in a car is __________________________.

The responses of 368 men to the 40 sentences were then subjected to a content analysis. (Content analysis was discussed earlier in the context of methods of measurement of attitudes and opinions, Chapter 11.) In this particular study, 98 specific categories of "content" were set up, and these, in turn, were grouped into 18 broader categories such as "desire for economy." By a special scoring scheme, each person was scored on each of these 18 categories. (These scores, in turn, were used in further analyses that need not be discussed here.)

---

[33] Jenifer S. MacLeod, *Predicting the Responses to Advertising Themes from Sentence Completions to Direct Attitude Questions About an Advertised Product* (New York: Advertising Research Foundation, Inc., 1958).

Construction techniques. In construction types of projective techniques, subjects are presented with a rather ambiguous type of stimulus (such as a picture) and are asked to create or "construct" a story, description, expression, or other form of response. Here, also, the intent is one of getting the subject to "project" himself in such a way that his attitudes or feelings may be revealed indirectly.

An interesting example of this technique as applied to advertising is reported by Engel and Wales,[34] in which cartoons were used to elicit responses. An example is shown in Fig. 19.4. This particular example dealt with the testing of possible themes for advertising of aspirin. In the cartoon, the druggist states (to a woman customer):

> This widely known brand of aspirin gives you 100 tablets for 67 cents, and this brand gives you 100 tablets for 27 cents. Which would you like?

The respondent is asked to "fill in" the answer of the customer. The responses can then be categorized such as by content analysis methods. In the survey in question, 150 housewives completed cartoon responses, and another 150 were asked the direct verbal question:

> Let's assume you needed some aspirin and went to a drug store to buy some. If you had your choice between a widely known brand of aspirin giving you 100 tablets for 67 cents and another brand giving you 100 tablets for 27 cents, which would you buy? Why?

Actually, the study was carried out in part to compare the responses obtained by these two methods. Some of the results are summarized below:

Response	*Cartoon respondents*	*Respondents to questions*
Expressed choice		
67-cent aspirin	48%	74%
27-cent aspirin	30%	18%
No choice	22%	8%
Mentioned quality	65%	40%
Mentioned need for further information	36%	7%

While the two procedures produced significantly different results, Engel expresses the opinion that the cartoon responses on brand choice may have provided the more accurate responses, since the respondents were under less implied pressure to show that they were concerned with aspirin "quality." Incidentally, this is another illustration of the possible influence of the method of data collection on the results obtained, and

[34] J. E. Engel and H. G. Wales, "Spoken *Versus* Pictured Questions to Taboo Topics," *Journal of Advertising Research*, 2 (March, 1962), 11–17.

J. E. Engel, *A Study of a Selected Projective Technique in Marketing Research*, Ph.D. Dissertation, University of Illinois, 1960.

Fig. 19.4. Illustration of a cartoon type of projective technique used in advertising research. The respondent is asked to fill in what she thinks the customer would say in response to the druggist's question. (From Engel and Wales, *op. cit.*)

therefore serves as a warning that such methods have to be chosen with considerable care, and the results evaluated accordingly.

Other Methods While the methods mentioned above are the more common ones in consumer research, there are other procedures that are used for some purposes. Perloff,[35] for example, refers to the use of various

[35] R. Perloff, "Determining Opinions and Attitudes through Survey Research," *Studies in Organization Management* (Washington, D. C.: Institute Department for Organization Management, Chamber of Commerce of the United States, 1962).

types of apparatus such as: eye cameras (such as used in recording eye movements of subjects when scanning advertisements, printed pages, television screens, etc.); mechanical devices for individuals to use in recording quickly their preferences of products or brands and of their opinions regarding packaging styles, etc.); and one-way mirrors. In addition there are observational techniques for observing consumer behavior in natural settings, either by having observers on hand (such as recording traffic patterns in a supermarket), or by use of photographic techniques.

In connection with taste testing, two rather common procedures are the use of triangle taste tests and paired preference tests. Triangle tests are used to determine if a particular food product can be differentiated by taste from another. For this purpose, two samples of one product and one sample of the other are given to the subject, with instructions to identify the one that is different from the others. The paired preference test involves samples of two related food products, with instructions to the subject to indicate which one he prefers.[36]

MARKETING RESEARCH

The various methods discussed above can be used in connection with different aspects of consumer research, one of the more important areas being market research. Of the various facets of marketing research as such, a few will be mentioned here.

Consumer Inclinations One aspect of marketing research relates to the tendency of consumers to be willing to buy, or not to buy, durable goods. This is of course a function both of the ability and willingness to buy. The "willingness," however, is, in part, a function of people's perceptions of economic prospects. Tendencies towards being optimistic versus pessimistic probably in part are associated with international and national affairs, as well as with information and impressions about such matters as unemployment, economic conditions, etc.

The Survey Research Center of the University of Michigan has for some years been measuring consumer attitudes and inclinations to buy, using an Index of Consumer Attributes developed by Katona.[37] Actually, two indices are used, one (based on six questions) relating to attitudes and expectations toward personal financial conditions, and market conditions, while the other (based on eight questions) also includes expressed intentions to buy. These two indices are shown in Fig. 19.5 for a period of several years.

---

[36] Some of the statistical aspects of triangle and paired tests are discussed by A. P. Radkins, "Some Statistical Considerations in Organoleptic Research: Triangle, Paired, Duo-Trio Tests," *Food Research,* 22 (No. 3, 1957), 259–265.

[37] G. Katona, "Consumer Attitudes and Inclinations to Buy in January–February, 1962," in *Psychological Research in Consumer Behavior* (Ann Arbor, Mich.: The Foundation for Research on Human Behavior, 1962).

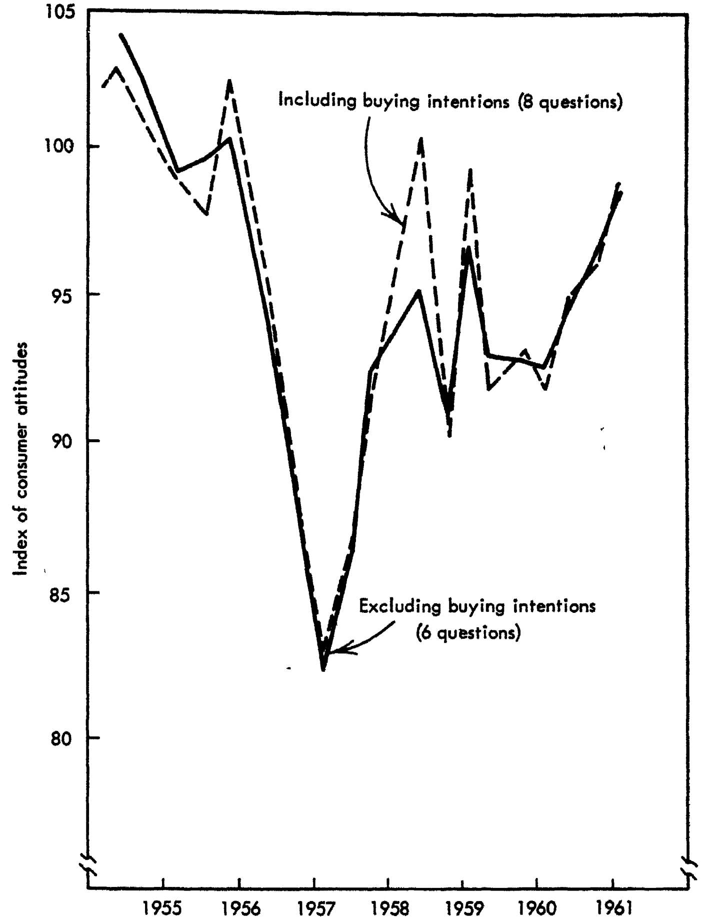


Fig. 19.5. Indices of consumer attitudes and inclinations to buy for a period of six years. (Adapted from Katona, *ibid.*, Table 2.)

Accompanying data relating to the expressed intentions of people to buy certain major items are shown in Fig. 19.6.

One might wonder, however, about the extent to which attitudes and expressed intent are related to subsequent buying behavior. More broadly, one might ask how accurately buying behavior can be predicted from various variables. An example of a study relating to this is reported by Mueller.[38] In the use of consumer surveys as a forecasting tool, her

38 E. Mueller, "Survey Methods as a Forecasting Tool," in *Psychological Research on Consumer Behavior* (Ann Arbor, Mich.: The Foundation for Research on Human Behavior, 1962).

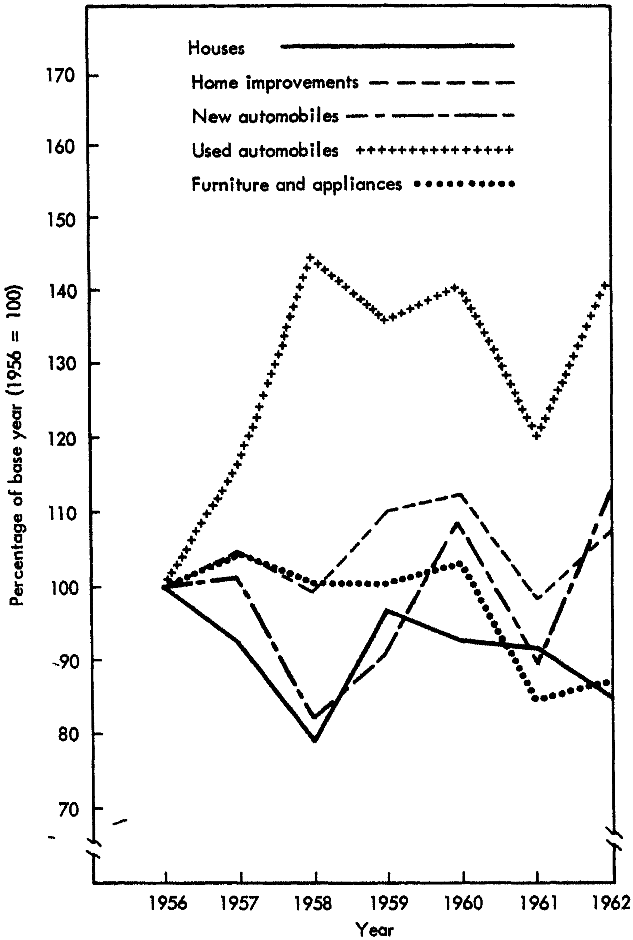


Fig. 19.6. Expressed intentions to purchase various major items, 1956–1962. (Adapted from Katona, *ibid.*, Table 3.)

basic premise is that discretionary spending by consumers is a function both of financial variables (consumer incomes, assets, debts, prices) and of attitudes and expectations. "Discretionary" expenditures generally refer to outlays for durable goods (cars, appliances, etc.), travel, recreation, and, to some extent, what one might think of as luxury foods and clothing. On the whole, consumer expenditures for "necessities" (food, shelter, etc.) are less influenced by financial variables or attitudes. In Mueller's [39] study she

[39] *Ibid.*

analyzed both financial and attitudinal variables as related to expenditures for durable goods as reflected by 21 surveys over a period of 9 years. Both types of variables were correlated with three measures of durable goods expenditures (for half-year following the survey in question), and multiple correlations were derived. Following are the weights of the variables that enter into three of the equations (multiple regression equations) to predict the three criteria, along with the multiple correlations (R) resulting from such equations as applied to the data from the 21 surveys:

Durable goods expenditures ($R = .88$)	$= 0.17(Y - 1) + 0.34A + 0.04B - 43.07$
Installment credit ($R = .81$)	$= 0.18(Y - 1) + 0.30A - 48.33$
New cars sold ($R = .80$)	$= 0.22Y_r + 0.06A_c - 22.01$
in which:	
$Y - 1$	= Disposable personal income during half-year preceding survey (Dept. of Commerce index)
Y_r	= Current disposable income as ratio to income 2 quarters earlier
A	= Survey Research Center Index of Consumer Attitudes (6 questions)
A_c	= Same as above, but modified to include 2 questions on attitudes toward automobile market
B	= Survey Research Center Index of Buying Attitudes

The three multiple correlations (R) of .88, .81, and .80 indicate that, to a very substantial degree, actual consumer expenditures for durable goods can be predicted on the basis of a combination of economic variables ($Y - 1$ and Y_r in the above equations) and attitudinal variables (A, A_c, and B in the equations).

Product Testing By pretesting products before they are placed on the market, manufacturers are able to avoid marketing products which the public will not buy in sufficient quantity to justify production and distribution costs. An example of this type of product testing has been published by Harris,[40] who discusses an approach to pretesting product designs and describes a study involving the validity of a method for predicting sales of dinnerware patterns. In one part of the study, ten fine china patterns were investigated. Ten dinner plates from each pattern being investigated were placed on a table and identified by a letter of the alphabet. Each of 140 women was brought individually to the table,

---

[40] Douglas Harris, "Predicting Consumer Reaction to Product Designs," *Journal of Advertising Research*, 4 (1964), 34–37.

instructed to assume that she was in the market for a set of fine china, and that the patterns before her were all available at the same price. She then ranked the patterns in order of her preference and indicated the degree of her preference for each pattern using a 21-point rating scale.

A Preference Index was then computed for each pattern, which was the percentage of respondents rating the pattern in the top three positions on the rating scale. The correlation between Preference Indices (obtained in this manner) and actual sales of the patterns (obtained from company records) was .91. This relationship is shown graphically in Fig. 19.7. It is obvious from Fig. 19.7 that patterns with high Preference Indices also are high in sales. Following this investigation, 70 new patterns were pre-

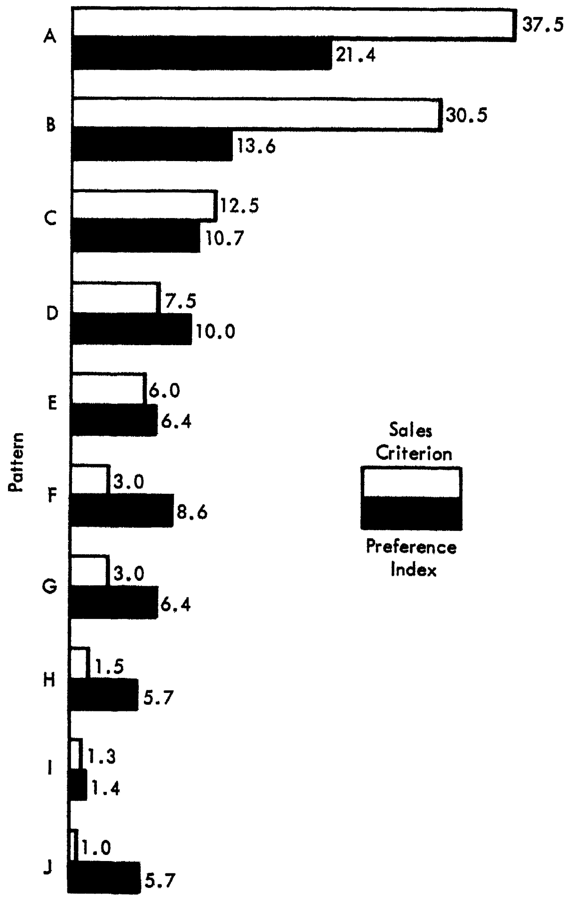


Fig. 19.7. Relation between Preference Index and Sales Criterion for ten fine china patterns (N-140). (From Harris, *op. cit.*)

tested on a national scale and 14 marketed on the basis of the results. Several patterns marketed on the basis of these pretest results have become top selling patterns in their grades.

Packaging Another aspect of marketing research is that relating to packaging of products. The producers of some products carry out research relating to package design in order to determine the opinions and preferences of people about various possible designs. Such opinions and preferences are usually obtained by the use of interviews with samples of customers. A rather ingenious method of obtaining responses is by the use of a pushbutton voting machine[41] that is installed in stores such as supermarkets. In one particular situation, two designs of packages of Vicks Vaporub were tested, one of which was the old design, the other a new design which shows a color reproduction of the familiar Vicks jar.[42] Three thousand machine votes were obtained in three days in selected supermarkets. The responses indicated quite clearly that the new design was the more popular.

Differential Characteristics of Consumers One of the important aspects of market research is related to the characteristics of people whose "behavior" as consumers is different. One study reported by Evans,[43] for example, dealt with some of the demographic (biographic) and personality variables that are associated with the choice of automobiles. Specifically the study dealt with a sample of Ford and Chevrolet owners in a Chicago suburb. The owners were all given a personality test (The Edwards Personal Preference Schedule) that measures ten personality "needs." In addition, certain biographical information was obtained about each individual. The statistical analyses were very much the same as in the validation predictors of job success (Chapters 4 and 5). Some of the results are given in Table 19.2. This table gives the average "scores" on each of the several items of biographical information for the two groups of owners. In addition, it gives the statistically determined weights that were used in deriving a multiple correlation of the several variables with brand ownership.[44] The most significant variables were: age of car; whether the owner smokes or not; whether the owner rents or not; having three or more children at home; and whether the owner had been with the same firm for five years or more. The weighting system shown in the table was used to compute a total score for each individual, and these scores, in turn, were used to predict which individuals owned Fords or Chevrolets. The scores predicted accurately in 70 per cent of the cases.

---

41 The mechanical testing machines are owned, installed, and serviced by the Automated Preference Testing Division of the A. C. Nielsen Co.

42 "A Robot Redesigns Vicks," *Modern Packaging*, 34 (No. 7, March, 1961), 90–91.

43 F. B. Evans, "Psychological and Objective Factors in the Prediction of Brand Choice," *Journal of Business of the University of Chicago*, 32 (October, 1959), 340–369.

44 The weights were determined by a linear discrimination function analysis

Table 19.2

AVERAGE "SCORES" OF DEMOGRAPHIC VARIABLES FOR FORD AND CHEVROLET OWNERS, AND STATISTICAL WEIGHTS ASSIGNED TO VARIABLES IN REGRESSION EQUATION *

Variable	*Scoring range*	*Ford (N=72)*	*Chevrolet (N=74)*	*Statistical weight*
Age of car	1 (1958) —4 (1955)	2.625	3.014	+1.0000 X_1
Over 10,000 miles per year	1 (over) —0 (under)	0.722	0.622	−1.0480 X_2
Shopped more than one dealer	1 (yes) —0 (no)	0.750	0.716	−0.1204 X_3
Owner smokes	1 (yes) —0 (no)	0.778	0.608	−2.1629 X_4
Own—rent	1 (Rent) —0 (Own)	0.417	0.554	+1.0189 X_5
Three or more children at home	1 (yes) —0 (no)	0.444	0.311	−0.8388 X_6
Catholic or not	1 (yes) —0 (no)	0.319	0.230	−3.4376 X_7
Protestant or not	1 (yes) —0 (no)	0.639	0.662	−2.8371 X_8
Attend church more than once a month	1 (no) —0 (yes)	0.375	0.460	+0.2189 X_9
Republican or not	1 (yes) —0 (no)	0.444	0.378	+0.3198 X_{10}
Democrat or not	1 (yes) —0 (no)	0.181	0.284	+1.7266 X_{11}
Age	1 (19) —9 (54)	5.333	5.351	−0.1304 X_{12}
Five or more years with same firm	1 (yes) —0 (no)	0.625	0.473	−1.0576 X_{13}
Income (mid-points)	1 ($3,750)—6 ($16,250)	3.194	3.068	+0.2482 X_{14}

* *From Evans,* op. cit., *Tables 6 and 7.*

It might be added that an analysis of the test scores indicated some differences. In brief, the Ford owners were found to have higher scores on the sub-tests of "need" for Exhibition and Dominance, and the Chevrolet owners on Autonomy and Affiliation.

The data revealed by a study such as this are, of course, highly specific to the situation. However, a study such as this demonstrates a quantitative approach to the analysis of the characteristics of different groups of consumers.

ADVERTISING EFFECTIVENESS

Another important aspect of consumer research is that relating to advertising effectiveness. It has been suggested by Lucas and Britt [45] that the measurement of advertisement effectiveness involves two different problems, namely, measuring advertisements and advertising effectiveness as such, and measuring the media which expose the advertising messages to consumers. These, however, are inseparably intertwined. While it is not

[45] Lucas and Britt, *op. cit.*, p. 3.

in order here to go into these in detail, examples of research relating to these two aspects will be given.

Advertising Research It has been suggested by Lucas and Britt[46] that there are four basic questions that need to be answered in the development of any advertising program, and four corresponding steps that are involved in an advertising copy research program. These are given below:

Question	*Research*
1. What do we know about the product or service?	1. Research prior to the development of the advertising
2. What shall we say?	2. Research during the development of the advertising
3. How shall we say it?	3. Pretesting of advertisements
4. How well did we do?	4. Post-testing of advertisements

In the evaluation of advertisement effectiveness, it has frequently been the practice to measure the effectiveness in terms of criteria of purchases of products by consumers. In this connection, however, it is frequently difficult, if not impossible, to trace down the effects on customer purchases of any given advertisement. The marketing "mix" is very complex; many factors intervene between seeing advertisements and buying products. In part because of this, it is becoming more common practice to use measures of attitude change as criteria of advertising effectiveness. As indicated earlier, attitudes are very important variables in all aspects of human behavior, including consumer behavior.

Media Research Advertisements can be presented by different media, such as newspapers, magazines, direct mail, outdoor billboards and car-cards, and television and radio. In comparing various media for communicating with potential consumers, one needs some measure of effectiveness. With respect to magazines (as an example), a very gross index is, of course, the number of copies sold. Such a measure, however, does not necessarily reflect readership. As indicated earlier, measures of recognition (also referred to as "noting") and of recall reflect different aspects of learning as the consequence of exposure (usually to particular articles or advertisements). Such measures have some relevance to the measurement of magazine audience. For example, the "audience" of magazines as characterized in one study were those who had seen at least one major editorial feature of the previous issue. Sometimes people are asked outright what magazines they read or prefer. There is, however, a potential pitfall to data obtained in such surveys. As pointed out by Lucas and Britt,[47] the act of reading, including that of reading most popular maga-

[46] *Ibid.*, pp. 10–11.
[47] Lucas and Britt, *op. cit.*, p. 223.

zines, carries with it a degree of prestige. This can lead to inflation of the number of people who say they read particular magazines. In one survey, for example, the number of people claiming to be readers of a given magazine was 15 times the number of copies printed! To avoid such inflation, the questions that are posed to the respondents (either by interviews or in questionnaires) should lead to no loss of status for admitting to the reading, or the failing to read, certain magazines. In this connection Lucas and Britt [48] point out that one technique that is used is that of having interviewers go through a magazine issue with a respondent, first noting the major editorial features, and afterwards ask such questions as: Just for the record, now that we have been through the issue, would you say you definitely happened to read it, or aren't you sure?

As an example of the results of media research, Table 19.3 shows the audiences for several magazines. In particular this shows the estimated total households, estimated total readers, the average circulation, and the estimated readers per copy. The readers per copy (of persons 18 years of age and over) is derived by dividing the estimated number of readers per copy (column 2) by the average circulation (column 3).

Table 19.3

AUDIENCES OF CERTAIN MAGAZINES IN THE UNITED STATES *
(000 OMITTED IN FIRST THREE COLUMNS)

Magazine	(1) *Estimated total households*	(2) *Estimated total readers*	(3) *Approximate average circulation*	(4) *Estimated readers per copy*
Esquire	4,030	4,830	779	6.2
Good Housekeeping	15,400	18,370	5,014	3.7
House Beautiful	5,540	6,530	972	6.7
Life	22,510	33,110	6,695	5.0
National Geographic	6,100	9,400	2,982	3.2
Reader's Digest	24,190	36,070	14,307	2.6
Sports Illustrated	6,380	7,670	1,023	7.5

* *Nielsen Media Service, 1963, copyright, A. C. Nielsen Company. Reproduced with permission.*

Where one is interested in "audiences" that are reached by different media, a question arises as to the total "unduplicated" audience that would be reached where some individuals may be exposed to two or more media, each of which carries the same advertisement. (If an ad appeared both in the *Saturday Evening Post* and in *Life,* those individuals who see both magazines should be counted only once in estimating total audience.) A method for estimating such unduplicated

[48] Lucas and Britt, *op. cit.,* pp. 228, 229.

audiences was reported by Agostini.[49] A modification of this method, reported by Kuhn [50] was tried out in comparison with empirically derived data for 180 combinations of magazines, with the result that maximum error was 1.8 per cent, with 130 of the combinations having errors of less than 1.0 per cent.

Television Ratings Somewhat related to media research are various ratings of television programs. One such set of ratings includes both a familiarity and a TvQ score, as defined below: [51]

Familiarity: The proportion of respondents with any opinion about a program.

TvQ score: A quantitative measurement of the degree of enthusiasm for a program. This score is determined by dividing the percentage saying that a show is "one of my favorites" by the percentage to whom the show is familiar. (It measures the intrinsic appeal of programs—not the size of the audience. It is a measure of preference.)

These indices are derived periodically from a sample of television viewers. For illustrative purposes, Table 19.4 gives these two indices for the nine most popular television shows (as identified here by letter).

Table 19.4

FAMILIARITY AND TVQ (PREFERENCE) SCORES OF NINE TELEVISION PROGRAMS

Rank	*Program*	*Familiarity*	*TvQ score*
1	a	84	55
2	b	88	48
3	c	89	46
4	d	74	43
5	e	71	43
6	f	59	42
7	g	80	41
8	h	81	40
9	i	58	39

Copyrighted by Home Testing Institute, Inc., 1962, Manhasset, N. Y.

Copy Testing It is frequently the practice in developing advertisements (especially for use in magazines with large circulations) to "test" the copy of alternative advertising designs. Usually this is done with samples of subjects before running the advertisements in any copies of the magazine. In some cases, however, different versions of advertise-

[49] J. M. Agostini, "A Short-Cut Method for Estimating the Unduplicated Audience of a Combination of Media," *Journal of Advertising Research,* March, 1961.

[50] W. Kuhn, "Net Audiences of German Magazines: A New Formula," *Journal of Advertising Research,* 3 (No. 1, March, 1963), 30–33.

[51] "TV's Favorite 15 Evening Programs as Measured by TvQ," *Media/scope,* 7 (No. 6, June, 1963), p. 96.

ments may be run in limited editions of the magazine (such as the copies prepared for different cities), and then tested by interview follow-up. In still other circumstances, each version of an advertisement may be included in all copies of a different issue of the magazine, and then tested. In such a case the purpose of the testing would be to provide guidance for the development of future advertisements.

An instance of this latter type is one in which two different advertisements for the Renault R-8 automobile were run in two separate issues of *Life* magazine, one in September, the other in November. These two versions were rated by both men and women, with their ratings being converted into Starch scores. Starch scores, developed by Daniel Starch & Staff,[52] provide the basis for a relative comparison of advertisements in terms of three types of criteria, namely "noted," "seen associated," and "read most," as based on the responses of subjects to interviewers' questions. In the case of these two advertisements, the results of these comparisons are given below:

	Starch Scores		
	Noted	*Seen associated*	*Read most*
Advertisement A			
Men	49	49	25
Women	19	15	7
Advertisement B			
Men	35	32	24
Women	12	7	4

It can be seen that there were generally higher ratings for advertisement A than for B, especially in the case of the "noted" and "seen associated" scores.

CONCLUSION

This chapter on consumer psychology has emphasized that methods of measurement are being used in this area as well as in the various other branches of applied psychology. The measurement may utilize interviews, quantified judgments, or statistical data coming from questionnaires or actual buying behavior.

Learning theory and research were discussed in connection with buying habits of people and also in measuring the effectiveness of advertising. It was shown how research methods make it possible to compare the relative effectiveness of space media, radio, and television in reaching (and influencing) a specified audience.

Methods of collecting data were discussed and ways to improve

[52] Daniel Starch & Staff, Mamaroneck, N. Y.

and evaluate the various methods were considered. In this connection, it was emphasized that the type of question used in a questionnaire often determines the answer one obtains. Different kinds of interviews, i.e., structured, unstructured, and depth were discussed. The advantages of each type in particular situations were then considered. Other methods of acquiring data that were summarized included projective and associative techniques, sentence completion tests, and related methods.

Marketing research was brought out as an important branch of consumer psychology. This area deals with consumer inclinations, product testing, packaging, and differential characteristics of consumers. Advertising effectiveness was then considered, and it was shown that quantitative methods are available to evaluate the relative effectiveness of advertisements appearing in different magazines. In this connection, methods of pretesting magazine advertisements before they are used were discussed. It was also pointed out that packages are often evaluated for appeal before being adopted by a company.

Throughout our discussion of consumer psychology, the concept of measurement was stressed. The methods of modern psychology make it possible to measure differences wherever differences exist. This is just as true in the area of consumer psychology as in the personnel psychology area.

Appendix A

Elementary Descriptive Statistics

DESCRIPTIVE *VS.* SAMPLING STATISTICS[1]

Two basic purposes are served by statistical methods. One of these is to describe a body of data. This is done by means of *descriptive statistics,* which deals with the reduction of a body of data by means of graphic methods, computational methods yielding numerical measures, and tabular methods. The sole aim of *descriptive statistics* is to reduce the original data to charts, graphs, averages, and the like so that the salient facts concerning the data will be more apparent.

The second purpose of statistical methods is to enable the experimenter to learn how safely he can generalize from *descriptive statistics* obtained on a sample. This approach is known as *sampling statistics.* It involves applications of the mathematics of probability and is a very important part of the field of industrial psychology.

This appendix will deal only with certain concepts of *descriptive statistics,* but the importance of *sampling statistics* should not be under-

[1] The distinction between *descriptive* and *sampling statistics* is carefully drawn by John G. Peatman, *Descriptive and Sampling Statistics* (New York: Harper & Row, Publishers, 1947). This book contains a clear and comprehensive treatment of both areas.

estimated and every serious student of industrial psychology should become familiar with these methods.

When many measurements, such as scores or other data, are to be summarized or interpreted, the use of some form of statistical procedure is usually desirable. If a considerable amount of raw data is involved, a simple listing of the data is of little value. Such a listing will not tell us, for example, how the data are distributed, how much they vary, or where in the total distribution they tend to cluster. Further, such a listing is of little value in indicating how the data compare with, or are related to, other sets of data collected under other circumstances. Before a meaningful interpretation of the data can be made it is necessary to reduce them to a *chart* or to *one or two single numbers that may represent the data as a whole.*

GRAPHIC REPRESENTATION OF DATA

The Frequency Distribution and Polygon A frequency polygon, constructed from a frequency distribution, is a graphic representation of a set of data. The construction and interpretation of a frequency polygon may best be explained by an example. Suppose 60 employees on an inspection job have detected the following number of flaws of a certain type during one week of work:

Table 20.1

NUMBER OF FLAWS DETECTED BY EACH OF 60 INSPECTORS DURING ONE WEEK OF WORK

15	36	40	37	32	13	35	20	33	36	33	16	38	19	33	34	24
36	25	29	27	39	42	31	21	26	28	53	23	51	21	26	39	28
30	31	32	30	29	49	39	30	44	34	37	35	38	35	41	37	43
42	38	45	22	46	41	47	48	34								

From a gross, or even a detailed, inspection of these 60 values one cannot answer such questions as: What is the typical number of defects spotted by an average inspector in a week? How much difference is there between the best and poorest inspectors in spotting defective materials? Is there any preponderance of good, poor, or average ability represented in the performance of these inspectors?

Such questions as these may be answered at a glance if the data are grouped and presented in a chart. One variety of such a chart is a frequency polygon. The steps involved in constructing a frequency polygon are as follows:

1. Determine the range of the values in the raw data. Quickly glance through the data to determine the *highest* and the *lowest* values. The range is the difference between these values. In the case of the 60

inspector records, the highest figure is 53 and the lowest is 13. The range is therefore 53 − 13 = 40.

2. If we find that the range of the data is larger (that they are widely spread), it will be more convenient to group them by intervals (class intervals, abbreviated c.i.) with a range in each c.i. of more than 1 unit. The c.i. is a group of adjacent scores of such a size that from 12 to 18 c.i.'s cover the range of the whole distribution or all of the data. With a range of 40, a c.i. of 2 would require 20 groups, a c.i. of 3, 14 groups; and a c.i. of 4, 10 groups. In our illustrative problem, a c.i of 3 is therefore the proper size to use.

A simple rule of thumb that is helpful in deciding upon the correct size of the c.i. is to divide the range by 15 (15 because, on the average, this is the most desirable number of c.i.'s) and take as the c.i. the whole number nearest to the quotient. In our problem, the range divided by 15 would be 40 ÷ 15 = 2.66. As 3 is the whole number nearest to 2.66, 3 would be the size of the c.i. to be used.

3. Arrange the adjacent c.i.'s in a column, leaving a blank space immediately to the right of this column. The arrangements of the c.i.'s preparatory to the construction of a frequency distribution appear as follows.

Table 20.2

CLASS INTERVALS TO BE USED FOR ILLUSTRATIVE DATA IN TABLE 20.1

51–53
48–50
45–47
42–44
39–41
36–38
33–35
30–32
27–29
24–26
21–23
18–20
15–17
12–14

4. Place a tally mark for each value in the original list of raw data opposite the appropriate class interval. As the first value among the 60 listed in Table 20.1 is 15, the first tally mark should be in the 15–17 c.i. The second value, 36, is represented by a tally mark in the 36–38 c.i. Usually it is advisable to tally the fifth entry in each c.i. with a line across the preceding 4 tally marks. This simplifies the counting of tally marks at a later time. When all entries have been made, that is, all data tabulated, the frequency distribution appears as in Table 20.3.

5. Lay off appropriate units on squared (cross section or graph) paper so that a graph may be constructed on which the midpoints of the c.i.'s are plotted on the base line and the frequencies or number of cases in each c.i. on the vertical axis. When this is done, the frequency polygon shown in Fig. 20.1 is obtained.

To one familiar with the concept of a frequency polygon, the graphic illustration in Fig. 20.1 is a much more meaningful repre-

Table 20.3

CLASS INTERVALS, TALLY MARKS, AND FREQUENCIES (F) FOR ILLUSTRATIVE DATA IN TABLE 20.1

Class intervals (*c.i.*)	*Tally marks*	*Frequency* (*f*)
51–53	//	2
48–50	//	2
45–47	///	3
42–44	////	4
39–41	~~////~~ /	6
36–38	~~////~~ ////	9
33–35	~~////~~ ////	9
30–32	~~////~~ //	7
27–29	~~////~~	5
24–26	////	4
21–23	////	4
18–20	//	2
15–17	//	2
12–14	/	1
		Total = 60

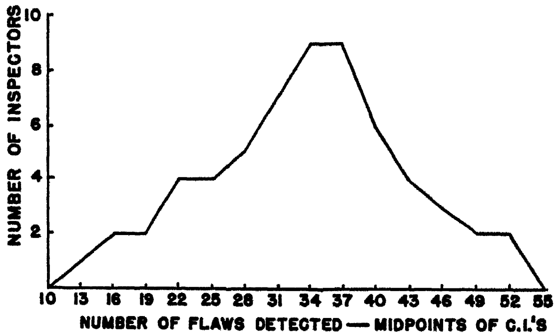


Fig. 20.1. Frequency polygon of illustrative data in Table 20.3.

sentation of the data than the list of values shown in Table 20.1, or the frequency distribution shown in Table 20.3. The frequency polygon makes apparent at a glance that the typical or average inspector de-

tected around 35 defects during a week of work, that the operators range or vary from some who detected only 13 defects to others who detected 53, and that a majority of the operators are fairly near the average in ability (that is, that not so many are very high or very low as are near the average). In summarizing psychological data it is a definite advantage to be able to present all of the major facts in a single graphic presentation of this type.

In a frequency polygon such as the one shown in Fig. 20.1, the area between the curve and the base line is determined by the number of cases (called N) the graph represents. Thus, a curve portraying 120 cases would cover twice the area of the curve shown (if the c.i.'s are the same for both distributions), and a curve portraying 600 cases would cover ten times this area. This is no disadvantage in many cases, but situations sometimes arise in which it is desirable to keep the total area under the curve the same, regardless of N or the number of cases. To meet this situation we may plot the *percentage* of cases rather than the *number* of cases falling in each c.i. This may be accomplished by dividing each f value in Table 20.3 by the total number of cases in the distribution. Each quotient thus obtained indicates the percentage of cases from the total falling in the respective c.i. These computations are indicated in Table 20.4.

If the percentages shown in the last column of Table 20.4 are now plotted as the ordinates (vertical axis points) of a frequency polygon, the chart shown in Fig. 20.2 is obtained.

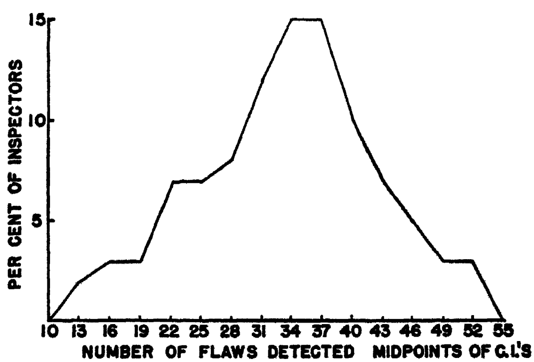


Fig. 20.2. Frequency polygon with ordinates as percentages for illustrative data in Table 20.4.

When a frequency polygon is to be compared with a number of other polygons, and when the important facts to be compared deal with the central tendencies and general form of the distributions rather than with the different number of cases plotted in each, the "percentage

method" of plotting a frequency distribution is preferred to the "total-number-of-cases method."

We may define a frequency polygon, in the light of the above description, as a *curve that portrays data graphically and that is so drawn that the base line represents the varying values of the original data and the ordinates represent the number of cases (or percentage of cases) at each of the raw data values.*

Table 20.4

FREQUENCY AND PERCENTAGE OF CASES IN EACH C.I. FOR ILLUSTRATIVE DATA IN TABLE 20 1

Class intervals	*f*	*Calculation of percentage*	*Percentage*
51–53	2	$2/60 = .033$	3
48–50	2	$2/60 = .033$	3
45–47	3	$3/60 = .050$	5
42–44	4	$4/60 = .066$	7
39–41	6	$6/60 = .100$	10
36–38	9	$9/60 = .150$	15
33–35	9	$9/60 = 150$	15
30–32	7	$7/60 = .117$	12
27–29	5	$5/60 = .083$	8
24–26	4	$4/60 = .066$	7
21–23	4	$4/60 = 066$	7
18–20	2	$2/60 = .033$	3
15–17	2	$2/60 = .033$	3
12–14	1	$1/60 = .017$	2
	Total = 60		100

The Normal Distribution The shape of the frequency polygons shown in Figs. 20.1 and 20.2 is typical of the kind of distribution usually found when data obtained from a group of people are plotted. It will be noted that the curves are approximately "bell-shaped," that is, they are high in the center and taper off toward the base line at both ends. If we were to divide the area under such a curve by drawing a perpendicular line from the central high point to the base line, the two parts would be approximately equal in area and would be bilaterally symmetrical in shape. It is well recognized that all, or nearly all, measurements of human traits and abilities result in distributions of approximately this form. Such distributions are called *normal distributions.* A strictly normal distribution conforms to a symmetrical bell-shaped curve that is defined by a mathematical equation, the basis of which is beyond the scope of the present discussion.[2] It will suffice for the beginning student to know that:

[2] *Ibid.*, 331 ff

1. A normal distribution is bell-shaped, that is, it is high in the center and low at both ends. Its two halves are symmetrical.
2. Measurements obtained from a group of persons usually approximate this type of distribution.

MEASURES OF CENTRAL TENDENCY

While the frequency polygon is helpful in giving an immediate graphic description of a set of data, that is, indicating general trends, it is often desirable to present certain quantitative figures that supplement the graphic picture. One of the most useful of such numerical values is a measure of the central tendency of the data. A measure of central tendency may be defined as a single figure or value that is representative of the entire set of data. Three such measures that are in common use are the arithmetic mean, the median, and the mode.

The Arithmetic Mean The arithmetic mean, sometimes simply called the mean, may be defined as the sum of the measures divided by the number of measures.[3] Or it may be thought of as a point of balance that could be found if all values in the distribution were assigned the same weight and then arranged along a horizontal beam. The physicist might define it as that point in the distribution around which the moments are equal.

In the case of the 60 values previously discussed from which a frequency polygon was constructed, the mean is obtained by finding the total of the 60 measures and dividing this total by 60, thus:

$$\text{Arithmetic Mean } (A.M.) = \frac{\text{Sum of measures}}{\text{N}} = \frac{2016}{60} = 33.6$$

This is the procedure followed in computing the exact value of the arithmetic mean of any set of values. In practice, a shorter method of computation utilizing data as tabulated in a frequency distribution and yielding an approximation (rather than the exact value) of the mean is often used. This shorter method assumes that each score as tabulated in a frequency distribution has the same value as the midpoint of the c.i. in which it falls. For further convenience in calculation, the mean is first computed in c.i. units from an arbitrary base selected near the center of the distribution at the midpoint of one of the c.i.'s. The base selected is entirely arbitrary—it may be taken as any point in the distribution. We have chosen one near the center of the distribution to simplify computation.

---

[3] *Ibid.*, 151 ff.

If this method is applied to the frequency distribution in Table 20.3 the arrangement shown in Table 20.5 is obtained.

Table 20.5

COMPUTATION OF THE ARITHMETIC MEAN FROM A FREQUENCY DISTRIBUTION

c.i.	*f*	*d*	*fd*
51–53	2	6	12
48–50	2	5	10
45–47	3	4	12
42–44	4	3	12
39–41	6	2	12
36–38	9	1	9
33–35	9	0	0
30–32	7	−1	−7
27–29	5	−2	−10
24–26	4	−3	−12
21–23	4	−4	−16
18–20	2	−5	−10
15–17	2	−6	−12
12–14	1	−7	−7
	60		$\Sigma fd = -7$

Formula for Computing *A.M.*

$A.M. = M^\circ + c.i.(c)$

M° = assumed mean

$c.i$ = size of $c.i.$

$c = \frac{\Sigma fd}{N}$ = summation of deviations from assumed mean divided by N

$$A.M. = 34 + 3\left(\frac{-7}{60}\right)$$
$$= 34 - 35 = 33.65$$

In the above tabulation the *d* column represents the number of c.i. units each c.i. is located above or below the c.i. arbitrarily chosen as the base for calculations. In the c.i. 51–53 there are two scores. This c.i. is 6 c.i. units above the arbitrary base. Thus, in computing the *A.M.* in c.i. units from the arbitrary base, these 2 scores would each have a value of 6, resulting in the number 12 that appears in the fourth, or *fd,* column. In like manner, there are 2 scores in the c.i. 48–50, and these 2 scores are each 5 c.i. units above the arbitrary base, resulting in the number 10 that appears in the *fd* column. All scores tabulated in c.i.'s below the arbitrary base are represented by negative values in the *fd* column. The algebraic sum of this column (Σfd) divided by the number of cases indicates how far the computed mean will deviate from the assumed mean (base) in terms of c.i. units. From the tabulation, this deviation in c.i. units from the arbitrary base (assumed mean) is defined as:

$$\text{Deviation in c.i. units from base} = \frac{\Sigma fd}{N}$$

Carrying through this computation for the data under consideration shows that:

$$\text{Deviation in c.i. units from base} = \frac{\Sigma fd}{N} = \frac{-7}{60} = -.117$$

This is interpreted to mean that the *A.M.* is .117 of a class interval below the midpoint of the arbitrary base (see formula in illustrative problem). In order to transmute this deviation (—.117) into raw score units, we would multiply it by 3 (the size of the class interval). Thus, in terms of raw score units, the deviation is −.35. The mean, as computed by this method, is therefore .35 raw score units below the midpoint of the 33–35 c.i. As the midpoint of this is 34, the mean is 34 − .35 = 33.65. This approximation does not agree exactly with the exact method in which all raw data are added and the sum is divided by the number of cases; but the approximation is sufficiently close to justify its use in many cases. The student may note, however, that essentially the same procedures are used in both solutions. The procedure in using the "exact method" may be thought of as involving the computation of a mean by assuming the mean to be zero, computing the deviations from zero in raw score units, and dividing their sum by N as in the short method.

The Median The median is a measure of central tendency defined as that score (or value) that exceeds, and is exceeded by, half the measures; that is, it is that point in the distribution above and below which 50 per cent of the values lie. A logical (though laborious) method to determine the median consists in arranging all the raw data in rank order from lowest to highest and counting off the bottom half of the measures. The value at this point is the median. If this method is followed for the data in Table 20.1, the following arrangement of the scores is obtained:

Table 20.6

ARRANGEMENT OF DATA FOR THE COMPUTATION OF THE MEDIAN DIRECTLY FROM RAW DATA

53	45	41	38	36	34	32	29	26	21
51	44	40	38	36	34	31	29	25	20
49	43	39	37	35	33	31	28	24	19
48	42	39	37	35	33	30	28	23	16
47	42	39	37	35	33	30	27	22	15
46	41	38	36	34	32	30	26	21	13

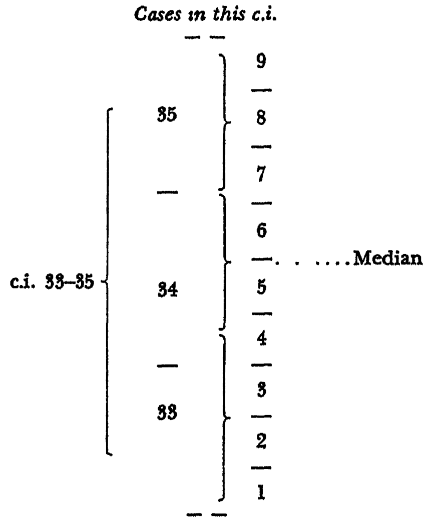
Counting from the lowest score up, we find that the thirtieth from the low end is 34, and the thirty-first from the low end is also 34. The median score would therefore be 34. If there had been a difference between the thirtieth and the thirty-first scores, the median would be the value halfway between these scores. If an odd number of cases were included in the original set of scores (as 61 instead of 60), the median would be the value of the middle score.

In practice, the median as well as the mean may be conveniently

approximated from a tabulated frequency distribution. To illustrate this process we may use the same frequency distribution previously discussed. The computations are shown in Table 20.7.

Knowing that 60 cases are included in the distribution, it is necessary to find the score that separates the lower 30 from the upper 30. The value of this score is the median. Counting up from the lower part of the distribution, we first fill the column *cumulative f,* which indicates for each c.i. the number of cases *in and below that c.i.*

It will be noted that 25 cases are included in or below the 30–32 c.i., and that 34 cases are included in or below the 33–35 c.i. The median, or point midway between the thirtieth and thirty-first cases, must therefore be within the 33–35 c.i. Now if we assume that the 9 cases in the 33–35 c.i. are distributed evenly throughout this interval, we must go up into this c.i. far enough to cover the lowest 5 of these 9 cases in order to reach the median. This may be illustrated graphically as follows:



It will be noted that the "real limits" of the 33–35 c.i. are considered 32.5 and 35.5; that is, the c.i. extends one-half a score unit above and one-half a score unit below the tabulated values. This is necessary because otherwise there would be a whole unit between each pair of adjacent c.i.'s that does not logically fall in either c.i.

In this case, the median would be five-ninths of the size of the c.i., added to the lower limit of the c.i., or Median $= \frac{5}{9}(3) + 32.5 = 34.2$. This value, 34.2, differs slightly from the value computed by arrang-

ing the scores in rank order, but the approximation is sufficiently close to justify its use in most instances. Exactly the same procedure can be followed in defining other points in the distribution, for example, the twenty-fifth percentile—the point in the distribution below which 25 per cent of the scores lie and above which 75 per cent lie.

Table 20.7

COMPUTATION OF THE MEDIAN FROM A FREQUENCY DISTRIBUTION

Class intervals	f	*Cumulative f*
51–53	2	60
48–50	2	58
45–47	3	56
42–44	4	53
39–41	6	49
36–38	9	43
33–35	9	34
30–32	7	25
27–29	5	18
24–26	4	13
21–23	4	9
18–20	2	5
15–17	2	3
12–14	1	1
Total	60	

Formula for Median

$$Md. = l + c.i. \frac{\left(\frac{N}{2} - F\right)}{f_m}$$

l = lower limit of $c.i.$ within which median lies.

$c.i.$ = class interval

$\frac{N}{2} = \frac{1}{2}$ of the scores.

F = no. of scores in all $c.i.$'s below l.

f_m = no. of scores in $c.i.$ in which median falls.

$$Md. = 32.5 + 3\,\frac{(60/2 - 25)}{9}$$

$$= 32.5 + 1.7 = 34.2$$

The Mode A third measure of central tendency is the mode, which is defined as the measure appearing most frequently. This value, as well as the mean and the median, may be determined directly from the raw data (if one value appears more often than any other) or may be approximated from a frequency distribution of the data.

In computing the mode directly from the raw data, the values are inspected to determine which one appears most frequently. Sometimes, as in the case of the values shown in Table 20.1, several of the measures appear an equal number of times. In this case, each of the values 30, 33, 34, 35, 36, 37, 38, 39 appears 3 times. It is incorrect, therefore, to say that any one of these is the mode. Furthermore, there is reason to believe that if a larger sample than 60 inspectors had been included in the distribution, and if the c.i. used in forming the distribution were smaller than 3, the frequency polygon obtained would be more even in curvature and only one single high point would be found. Under such circumstances, this high point would be the mode. An approximation of

this value may be obtained from a frequency distribution by means of the following empirical formula:

$$\text{Mode} = 3(\text{Median}) - 2(\text{Mean})$$

In the case of the data we have been discussing, this formula gives the following value for the mode:

$$\text{Mode} = 3(34.2) - 2(33.65) = 35.30.$$

When to Use the Mean, Median, and Mode Why is it necessary to have 3 different measures to indicate the central tendency of a set of data? The answer is that each is best adapted to certain uses; that is, in some cases one may be most representative of a set of data, while in other cases another measure may be most suitable. The mean is ordinarily used if the distribution is approximately normal. (If the distribution is perfectly normal, the 3 measures of central tendency have the same value.) If, on the other hand, there is a preponderance of extreme cases at either end of the distribution, the mean may give an incorrect impression of the central tendency of the data. Under these circumstances, the median or mode is more suitable. Consider, for example, the following yearly incomes of five persons:

$800 $900 $850 $750 $5,000

The mean for these five incomes is

$$\frac{\$800 + \$900 + \$850 + \$750 + \$5{,}000}{5} = \$1{,}660.$$

This figure, though an accurate statement of the mean, is not typical of the group as a whole because it is so markedly affected by the one income of $5,000 that is considerably larger than the other four. The median income is $850, and this value is more typical for the group as a whole than is the mean income of $1,660. If a great deal of data were available for computation, it would also be enlightening to know the mode, or most common income. Certain other principles also help determine which measure of central tendency is most appropriate in any specific case. We may generalize in the above illustration by saying that if a distribution is very much *skewed* (that is, contains more cases at one extreme than at the other), the median or mode is more likely to give a representative picture of the typical score than is the mean.

MEASURES OF VARIABILITY

In addition to a measure or value to represent the central tendency of a set of data, there is also quite frequently a need for some measure

of the spread, or variability, of the data. The need for a measurement of this type may be seen by comparing the data shown in Table 20.1 and tabulated in Table 20.5 (the mean of which, computed from the frequency distribution, was found to be 33.65) with another set of data that, for purposes of illustration, we might assume to consist of 21 scores of 33, and 39 scores of 34, making 60 scores in all. The mean of 60 such scores may readily be found to be 33.65.

$$\frac{(21)(33) + (39)(34)}{60} = 33.65$$

While both distributions have the same mean, they differ markedly in variability or spread. The former distribution is made up of scores varying from 13 to 53, while the latter consists entirely of scores of 33 and 34. A quantitative measure of variability is therefore of considerable value. Statistical procedures have been designed that yield a single value descriptive of this variability; as in the cases of means and medians these measures tell us something about the group as a whole.

The Standard Deviation The Standard Deviation is the most widely used measure of variability. It is defined as the square root of the mean square deviation. Defined by formula:

$$\text{Standard Deviation} = S.D. = \sigma = \sqrt{\frac{\Sigma D^2}{\mathrm{N}}}$$

where ΣD^2 is read "the sum of the squared deviation of the scores from their mean" and N is the number of cases. S.D. and σ are abbreviations for the Standard Deviation. They are used interchangeably.

Although the Standard Deviation may be computed directly from a set of raw data by means of the formula $S.D. = \sqrt{\frac{\Sigma D^2}{\mathrm{N}}}$, this process is laborious. For example, in the case of the set of data we have been using for illustrative purposes (tabulated in Table 20.1), we would proceed by determining the difference between each raw score and the mean of the 60 scores, squaring these differences, summing the 60 squared differences, dividing by 60, and extracting the square root of the quotient. The first score tabulated is 15. The difference between this value and the mean of the 60 scores (as computed directly from the raw data) is $D = 33.6 - 15.0 = 18.6$. D^2 would therefore be $(18.6)^2 = 345.96$. This must be repeated for every one of the 60 scores before the sum of the squared deviations can be obtained.

Because of the excessive labor in computing the S.D. directly from the raw data, a simple process that approximates the true value of the S.D. has been developed. This is used in the computations shown in Table 20.8, in which is computed the Standard Deviation of the data shown previously in Table 20.5.

Table 20.8

COMPUTATION OF THE STANDARD DEVIATION FROM A FREQUENCY DISTRIBUTION *

c.i.	*f*	*d*	*fd*	*fd²*
51–53	2	6	12	72
48–50	2	5	10	50
45–47	3	4	12	48
42–44	4	3	12	36
39–41	6	2	12	24
36–38	9	1	9	9
33–35	9	0	0	0
30–32	7	−1	−7	7
27–29	5	−2	−10	20
24–26	4	−3	−12	36
21–23	4	−4	−16	64
18–20	2	−5	−10	50
15–17	2	−6	−12	72
12–14	1	−7	−7	49
Total = 60			$\Sigma fd = -7$	$\Sigma fd^2 = 537$

$$\text{Mean} = M^\circ + c.i\left(\frac{\Sigma fd}{\text{N}}\right)$$

$$\text{Mean} = 34 + 3\left(\frac{-7}{60}\right) = 33.65$$

$$\text{S.D.} = \sqrt{\frac{\Sigma D^2}{\text{N}}} = c.i.\sqrt{\frac{\Sigma fd^2}{\text{N}} - \left(\frac{\Sigma fd}{\text{N}}\right)^2}$$

$$= 3\sqrt{537/60 - (.117)^2}$$

$$= 3\sqrt{8.950 - 014}$$

$$= 3\sqrt{8.936}$$

$$= 3(2.99)$$

$$\text{S.D.} = 8.97$$

* The derivation of the formula for computing the standard deviation by the method used in Table 20.8 is given below. Working in c.i. units rather than raw score units, let

$D_1, D_2, \ldots D_n$ = deviations of the scores in c.i. units from their mean.
c = difference in c.i. units between the mean of the scores and the midpoint of zero c.i.; i.e. the arbitrary base.
$d_1, d_2, \ldots d_n$ = deviations of the scores in c.i. units from arbitrary base.

Then

$$D_1 = d_1 - c$$
$$D_2 = d_2 - c$$
$$\cdots\cdots\cdots$$
$$D_n = d_n - c$$

If the above equations are squared on both sides, we have

$$D_1^2 = (d_1 - c)^2 = d_1^2 - 2d_1c + c^2$$
$$D_2^2 = (d_2 - c)^2 = d_2^2 - 2d_2c + c^2$$
$$\cdots\cdots\cdots\cdots$$
$$D_n^2 = (d_n - c)^2 = d_n^2 - 2d_nc + c^2$$

Summating the above to determine the sum of the squared differences, we have

$$D_1^2 + D_2^2 + \ldots D_n^2 = (d_1^2 + d_2^2 + \ldots d_n^2) - (2d_1c + 2d_2c + \ldots 2d_nc) + \text{N}c^2$$

The above, written with summation signs, becomes:

$$\Sigma D^2 = \Sigma d^2 - 2c\Sigma d + \text{N}c^2 \tag{1}$$

The Standard Deviation is the most commonly used measure of variability. Usually when the mean value of a set of data is given, the S.D. is also given to indicate the variability of the data.

COMPARABLE SCORES

The S.D. performs another useful function—it can be used in comparing individual scores from different distributions. For example, suppose that 2 inspectors from departments *A* and *B*, who are working at different inspection jobs, detect respectively 45 and 89 defects during a week of work. How can we compare the efficiency of these 2 employees? It will be seen immediately that a direct comparison of the figures 45 and

---

Now it will be remembered from Table 20.5 that c, which is the difference in c.i. units between the mean of the data and the arbitrary base (i.e. the mean in c.i. units away from the arbitrary base), was determined by adding the d values of the scores and dividing this sum by the number of scores. Since there are f_1 scores at d_1 deviation; f_2 scores at d_2 deviation; etc., this summation is given by $f_1d_1 + f_2d_2 + \ldots f_nd_n = \Sigma fd$, which, divided by N, gives $\frac{\Sigma fd}{N}$.

We may therefore substitute in (1) $\frac{\Sigma fd}{N}$ for c, giving

$$\Sigma D^2 = \Sigma d^2 - 2\frac{\Sigma fd}{N}\Sigma d + N\left(\frac{\Sigma fd}{N}\right)^2$$

But the Σd and Σfd are the same, since Σfd is only a simpler method of determining the Σd that involves grouping together all scores of the same d, and multiplying this d by f, the number of such scores. For the same reason, Σd^2 is the same as Σfd^2. The above equation therefore may be written:

$$\Sigma D^2 = \Sigma fd^2 - 2\frac{\Sigma fd\,\Sigma fd}{N} + N\left(\frac{\Sigma fd}{N}\right)^2$$

$$\Sigma D^2 = \Sigma fd^2 - 2\frac{(\Sigma fd)^2}{N} + \frac{(\Sigma fd)^2}{N}$$

If both sides of the above equation are divided by N, we have

$$\frac{\Sigma D^2}{N} = \frac{\Sigma fd^2}{N} - 2\left(\frac{\Sigma fd}{N}\right)^2 + \left(\frac{\Sigma fd}{N}\right)^2$$

$$\frac{\Sigma D^2}{N} = \frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2$$

Extracting the square root of both sides

$$S.D. = \sqrt{\frac{\Sigma D^2}{N}} = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$$

which is the equation used in the computations accompanying Table 20.8 to determine the standard deviation of the distribution in c.i. units. The value yielded by this expression is then multiplied by the size of the c.i. to give the *S.D* in raw score units.

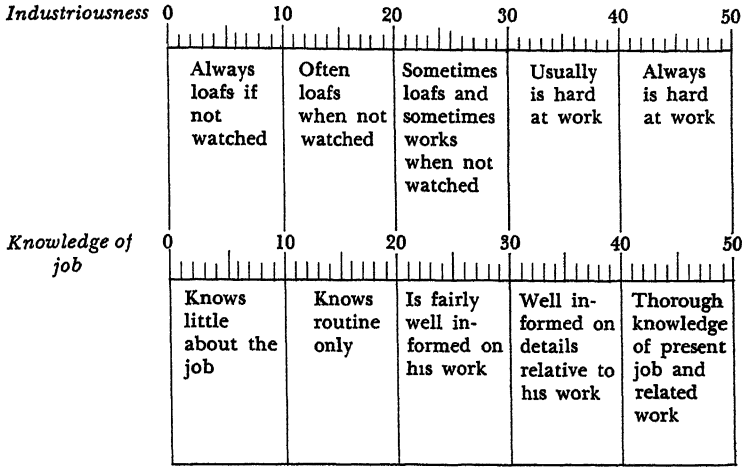
89 is not valid, because the 2 inspection jobs may be very different. It will also be seen that we can say little concerning the position of these inspectors in their respective groups without knowing their relation to the mean of their group in inspection work. To make a comparison, then, we must first compute the mean number of defects spotted by all inspectors in Department *A*, and the mean number spotted by all inspectors in Department *B*. Suppose that those means are respectively 38 and 95. We thus see that the inspector from Department *A* is $45 - 38 = 7$ pieces *above* the mean for that department, and that the inspector from Department *B* is $89 - 95 = -6$, or 6 pieces *below* the mean of inspectors from that department. We can thus say, at this point, that the inspector from Department *A* is above average in ability on the job and that the inspector from Department *B* is below average. But how about their relative distance from the average? To answer this question we must compute the S.D.'s of the 2 distributions and determine how many S.D.'s each inspector is above or below average.

Suppose we find the S.D. of the operators in Department *A* to be 5.5 pieces. Our first inspector is therefore $\frac{45 - 38}{5.5} = 1.27$ S.D.'s above average. If the S.D. of the inspectors in Department *B* is 9.5, the inspector from the group who detected 89 pieces is $\frac{89 - 95}{9.5} = -.63$ or .63 S.D.'s below average. *The deviation of a score from the mean of the distribution expressed in S.D. units results in a measurement that is comparable with similarly determined measurements from other distributions.* Thus, we may say that our first inspector is about *twice* as far above average, in terms of comparable scale units, as the second operator is below average. Scores computed in this manner are known as *z*-scores. The formula for a *z*-score is as follows:

$$z\text{-score} = \frac{\text{Raw Score} - \text{Mean of Raw Scores}}{S.D. \text{ of Raw Scores}}$$

The *z*-score is helpful not only when comparing scores from one distribution to another but also when, for any reason, it is desired to combine scores with the same or differential weighting. A typical example of an industrial situation that requires this technique is in the combination of items used in a merit-rating blank. Suppose that each employee has been rated by his supervisor on a chart containing items such as the one on p. 631.

We may suppose, for purposes of illustration, that it is now desired to combine these 2 traits into an over-all merit rating. (If more than 2 traits are included in the chart, as is usually the case, the procedure is identical.) Suppose that an employee, Mr. A, has received 40 points on industriousness and 30 points on knowledge of job, making a total of 70 points if the ratings are added directly. Suppose that another employee, Mr. B, has received 30 points on industriousness and 40 points on knowl-



edge of job, which also results in a total of 70 points if added directly. It is clear that such direct and immediate combination of ratings would result in identical over-all ratings for these 2 employees. The question we may raise is whether such a statement of equal ratings is justified. The answer is that it is not. If the mean rating of all employees on industriousness was 33 with a S.D. of 3, then A's rating would be $\frac{40 - 33}{3}$ $= 2.33$, or 2.33 S.D.'s above the mean, and B's would be $\frac{30 - 33}{3} = -1.00$, or 1.00 S.D. below the mean on this trait. If the mean rating for all employees on knowledge of job were 25, with a S.D. of 6, A would be $\frac{30 - 25}{6}$ $= +.83$, or .83 S.D.'s above average in knowledge of job while B would be $\frac{40 - 25}{6} = 2.50$ or 2.50 S.D.'s above average in this respect. Now, the proper combination of the 2 traits, if we wish to weight them equally, would be:

Employee	*Rating in industriousness*	*Rating in knowledge of job*	*z-Score in industriousness*	*z-Score in knowledge of job*	*Sum of scores for both units*
A	40	30	+2.33	+.83	+3.16
B	30	40	−1.00	+2 50	+1.50

This transfer of ratings into *z*-scores and the adding of the *z*-scores show that the 2 employees A and B are not equal in rating (as we would infer if the raw ratings were added), but rather that A is definitely higher than B. The procedure described has assumed that the 2 trait ratings, being combined, should be given equal weight, and the procedure shows

how they can be combined with equal weight into a composite score. One might think that conversion of raw scores to *z*-scores is not necessary if the raw scores are to be given equal weight in the combination score. Actually, if we do not give the raw scores equal weight by converting them into *z*-scores, the scores will weight themselves according to the size of their respective Standard Deviations. In other words, if combined directly, the raw scores will be weighted too much or too little, depending upon their position relative to the means of their respective distributions and upon the variability of the distribution of which they are a part. When scores are combined, they are *always* weighted in some manner, whether we deliberately weight them or not. It is highly important, therefore, to weight them deliberately (either with equal weight or otherwise) by converting them into *z*-scores and then combining them.

It does not follow from the above discussion that combined scores should always be weighted equally. Indeed, it is often desirable to weight various scores according to some plan that has been decided upon before the scores are combined. When this is desired, such weighting can be accomplished very easily by multiplying each *z*-score by the appropriate weight before they are combined. In our illustrative case, suppose that we have decided that *industriousness* should be given twice as much weight as *knowledge of job* in determining the total rating. This would be accomplished as follows:

Employee	*z-Score in industriousness*	*z-Score in knowledge of job*	*Weighted z-score in industriousness*	*Weighted z-score in knowledge of job*	*Combined weighted z-scores*
A	+2.33	+.83	+4.66	+.83	+5.49
B	−1.00	+2.50	−2 00	+2.50	+.50

The combined ratings so obtained show a still greater difference between employees A and B than was obtained when the scores were equally weighted. If, on the other hand, it is desired to give the rating on *knowledge of job* twice as much weight as the rating on *industriousness,* the following computations would be made:

Employee	*z-Score in industriousness*	*z-Score in knowledge of job*	*Weighted z-score in in industriousness*	*Weighted z-score in knowledge of job*	*Combined weighted z-scores*
A	+2 333	+ 833	+2.333	+1 666	+4.0
B	−1.000	+2.500	−1 000	+5.000	+4.0

This last procedure results in giving identical total scores to employees A and B, which, it will be remembered, also occurred when the

raw scores were added directly (30 + 40 = 70). Now, if we remember that the assumed Standard Deviation for the *knowledge-of-job* ratings was 6, which is twice as great as the Standard Deviation of 3 assumed for the *industriousness* ratings, we can see why a direct combination of raw scores gives the equality that is obtained when the respective *z*-scores are weighted in the ratio 2:1. In this instance the original ratings whose Standard Deviation is 6 are automatically given twice as much weight (when raw scores are combined directly) as the original ratings whose Standard Deviation is only 3. Thus, we obtain the same final result (equality) by direct combination that is obtained when the *knowledge-of-job* ratings are deliberately given a weight twice as great as the *industriousness* ratings.

Many other problems arise in which it is necessary to weight scores to achieve a particular result. For example, a company faced with the problem of selecting a number of electrical apprentices desires to give this training to those boys who have the greatest aptitude for the job and who are therefore most likely to succeed. Careful consideration and discussion of the problem by management and supervision resulted in the decision that four factors should determine whether an employee should be given this training. These 4 factors were general intelligence, present knowledge of electricity, previous merit rating, and seniority with the company. It was further decided in conference that, although all of these four factors should be considered, they are not of equal importance. It was decided that a fair weighting of their relative importance was as follows:

General intelligence	40%
Knowledge of electricity	30%
Merit rating	20%
Seniority or service with the company	10%

To score the applicants according to this plan, each was given a general intelligence test and a test covering technical phases of electricity. Merit ratings and seniority were obtained from the company records. Each of the four scores was converted into a *z*-score and the four resulting *z*-scores were respectively multiplied by 40, 30, 20, and 10. For each employee the sum of the weighted *z*-scores was used in indicating whether or not he was given the apprenticeship training.

PERCENTILES

The discussion of comparable scores should have made clear the fact that a raw score on any test is relatively meaningless unless it is interpreted in terms of its location in a distribution of other scores made by other people. If a test consists of 75 very easy questions, a score of 65

might be near the bottom of the distribution and hence should be interpreted as a very low score. On the other hand, if a test consists of 75 very difficult questions, a score of 65 might be at or near the top of the distribution and should therefore be considered a very high score. In other words, a raw score of 65 might be a low score or a high score, depending upon the distribution of scores from which it is drawn.

One convenient and widely used method of interpreting a raw score is by using *percentile ranks.* A percentile rank may be defined as the number showing the percentage of the total group equal to or below the score in question. Thus, on a certain test, if 65 per cent of the total group scored 129 or below, the score of 129 would be at the 65th percentile, or would have a percentile rank of 65. The 50th percentile, it will be noted, is the same as the median as previously defined.

A convenient, practical method of determining by close approximation the percentile equivalents of a set of raw scores makes use of a cumulative frequency distribution such as the one tabulated in Table 20.9. This tabulation is based on the same distribution previously used in Table 20.3.

Table 20.9

CUMULATIVE FREQUENCY DISTRIBUTION USED IN DETERMINING PERCENTILE RANKS OF RAW SCORES

(1) *Class intervals*	(2) *f*	(3) *Cumulative f*	(4) *Per cent*
51–53	2	60	100
48–50	2	58	96
45–47	3	56	93
42–44	4	53	88
39–41	6	49	81
36–38	9	43	71
33–35	9	34	56
30–32	7	25	42
27–29	5	18	30
24–26	4	13	22
21–23	4	9	15
18–20	2	5	8
15–17	2	3	5
12–14	1	1	2
	Total = 60		

The per cent values, in column (4) of Table 20.9, are obtained by dividing each of the values in the cumulative *f* column, column (3), by the total of column (2), in this instance 60. The per cent values in column (4) are then plotted against the upper limits of the class intervals, as shown in Fig. 20.3. From Fig. 20.3 the percentile ranks of the raw scores may be read directly, with sufficient accuracy for most purposes.

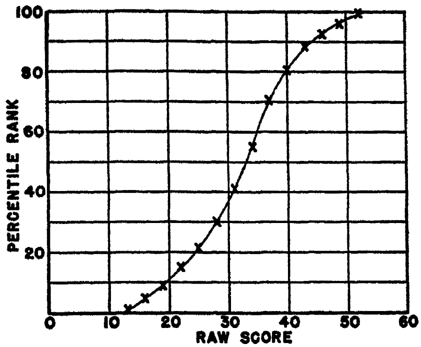


Fig. 20.3. Chart for converting raw scores into percentile ranks for illustrative data in Table 20.10.

When this has been done, the raw scores with corresponding percentile ranks may be tabulated as illustrated in Table 20.10.

In Table 20.10, only certain percentile ranks are given, because these are usually sufficient for ordinary purposes. The intervening percentile ranks, however, may also be read from Fig. 20.3, if this is necessary.

The manual published with standardized tests usually includes percentile tables making possible the conversion of raw scores into percentile ranks. Since the percentile rank of a given raw score is dependent upon the nature of the group used in constructing the conversion table, several raw-score-to-percentile-rank conversion tables, based on different groups, are often published with standardized tests.

Table 20.10

RAW SCORES AND PERCENTILE RANK EQUIVALENT READ FROM FIG. 20.3

Percentile rank	*Raw score*
100	53
98	52
95	48
90	44
80	40
70	37
60	35
50	33
40	31
30	28
20	24
10	20
5	17
2	14
1	13

CORRELATION

In numerous experimental situations 2 variable quantities are so related that they vary, or tend to vary, with each other. A common problem in industrial psychology is to reduce to a simple and meaningful statement the facts that have been discovered concerning such a functional relationship. Suppose that a number of punch-press operators during a given period of time have each punched a certain number of pieces and have each mispunched, or otherwise wasted, a certain number of pounds of stock material. In such a situation, it might be of considerable importance for management to know whether any relationship exists (and, if so, how much) between quantity of work done and amount of material wasted. Indeed, the company's policy with respect to speed of work recommended as well as the quality control in the form of penalty or bonus might well be formulated correctly only in the light of specific knowledge of the relationship between speed and accuracy.

Consider a department employing 8 operators for whom the following figures for production and waste in pounds are available:

Operator	*Production*	*Waste*
1	95	3.0
2	103	4.5
3	88	3.5
4	98	4 0
5	93	3.0
6	107	4.5
7	114	4.0
8	106	5.0

It is difficult, if not impossible, to determine from a gross inspection of these 2 columns of figures whether any relationship exists between speed and waste. It is necessary to employ some type of graphic or computational procedure to determine the amount of relationship that may exist between these 2 sets of data. One simple and sometimes satisfactory method consists of a simple plotting of the values on co-ordinate axes and rough inspection of the results. If we let production be represented on the x or horizontal axis, and waste on the y or vertical axis, then the production and waste of each operator will locate him on a chart, giving the result shown in Fig. 20.4.

A plot such as is shown in Fig. 20.4 gives a much better indication of the presence or absence of a relationship between the data than can be obtained from the columns of raw data from which the chart was prepared. The chart shows that some relationship does exist, and it is

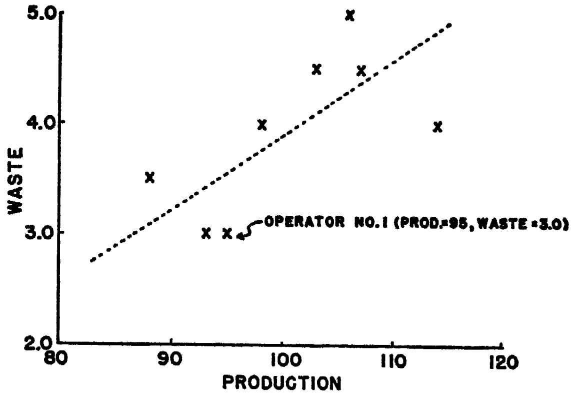


Fig. 20.4. A plot of the production and waste records for the eight punch-press operators shown above.

even possible to draw in by inspection a line or curve that represents this relationship in an approximate form.

Although this simple method of studying the relationship between 2 variables is sometimes adequate for very simple problems or for those that involve only a small amount of data, it is not adequate for an exact study because it does not result in a quantitative statement of the degree of relationship. The slope of the dotted line cannot be considered such a quantitative statement because: (1) this line is drawn in by inspection, and (2) its slope depends upon the units of measurement on both the x and y axes.

Two commonly used quantitative methods for measuring the degree of relationship between 2 paired sets of data are rank-order correlation and product-moment correlation.

Rank-Order Correlation The use of this method may be described by applying it to the data for the 8 punch-press operators.

In Table 20.11 the 2 columns headed *Rank* give, respectively, the rank of the operators on the 2 measures (production and waste). The highest producing operator (in this case the seventh in the list) is given a rank of 1, the second highest a rank of 2, and so on. In like manner, the rank of each operator in wastage is placed in the waste-rank column. In case 2 or more operators are tied for a given rank (as in the case of the second and sixth operators, who are tied at 4.5 pounds of waste each), the tied scores are all given the same rank, which is the average of the ranks that would have been assigned to the tied scores if they had not been tied. The values in the D^2 column are obtained by squaring each D

value. The sum of the D^2 column is determined and the correlation computed by means of the formula:

$$\text{Rho} = 1 - \frac{6\Sigma D^2}{N(N^2 - 1)} \cdot$$

N is the number of cases entering into the computation.

This formula for the rank-order correlation is an empirical formula. It yields a value of +1.00 if the data are in exactly the same rank order. (The reason for this may be seen from the fact that if all ranks are the same, all D's are zero, all D^2 values are zero, ΣD^2 is zero, and the formula becomes $1 - 0 = 1$.) If the data are in exactly reverse order (that is, if the individual who ranks highest on one series is lowest on the other, and so on), the formula will yield a value of —1.00, but if no relationship exists between the two sets of data, a correlation of zero will be found.

Table 20.11

COMPUTATION OF RANK-ORDER CORRELATION

Operator	*Production*	*Waste*	*Rank in production*	*Rank in waste*	*Difference in rank (D)*	*(D)²*
1	95	3.0	6	7 5	1.5	2.25
2	103	4.5	4	2.5	1.5	2.25
3	88	3.5	8	6.0	2 0	4.00
4	98	4.0	5	4 5	.5	.25
5	93	3 0	7	7.5	.5	.25
6	107	4.5	2	2.5	.5	.25
7	114	4 0	1	2.5	3.5	12.25
8	106	5.0	3	1.0	2.0	4.00
						25.50

$$\text{Rho} = 1 - \frac{6\Sigma D^2}{N(N^2 - 1)} = 1 - \frac{153}{504} = .70$$

The use of this formula is ordinarily more satisfactory than a simple plotting of one variable against the other because it yields a quantitative statement of the degree of relationship and not simply a graphic representation that cannot be reduced to a numerical statement.

If an appreciable number of cases are involved, however, the rank-order method of computing the degree of relationship is extremely laborious. For this reason—and for other reasons of a mathematical nature—it is ordinarily used only when the data are limited to a very few cases (less than 30).

The Product-Moment Coefficient This is the most widely used measure of relationship. Like the rank-order correlation, it may vary from +1.00 (indicating perfect positive relationship) through zero (indicating

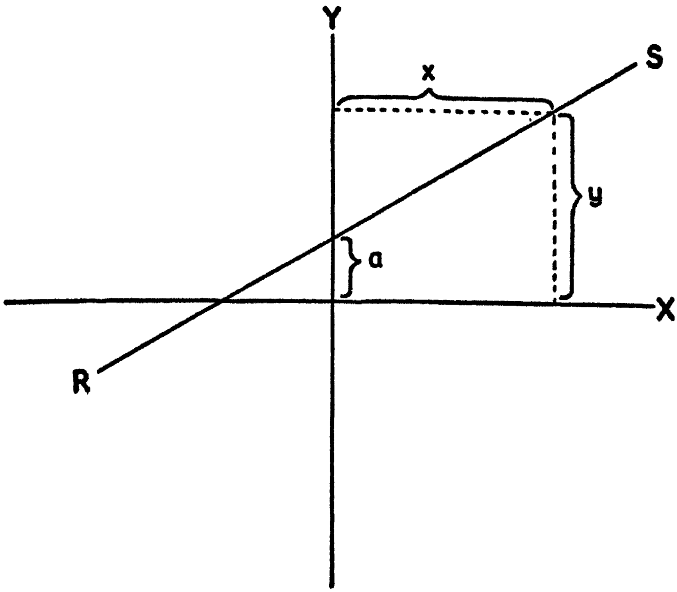


Fig. 20.5. The slope of the line RS is defined as $b = \frac{y-a}{x}$.

no relationship) to −1.00 (indicating perfect negative relationship). The product-moment correlation, represented by the symbol r, may be defined in several ways. One of the simplest definitions is that *r is the slope of the straight line that best fits the data after the data have been plotted as z-scores on co-ordinate axes;* that is, it is the tangent of the angle made by this line with the base line.

Several terms in this definition require further definition. By *slope* is meant steepness with which the line rises. The slope of a straight line drawn in any manner across co-ordinate paper is defined as the distance, y, from any given point on the line to the x intercept, minus the distance, a, from the origin to the y intercept, divided by the distance, x, from the point on the line to y intercept. Thus the slope, which we will call b, is defined in Fig. 20.5 as follows:

$$b = \frac{y - a}{x}$$

It should be remembered that, on co-ordinate axes, distances above and to the right of the origin are positive, while distances measured below and/or to the left of the origin are negative. The slope of any line that *rises* as it goes from left to right will therefore be positive (the greater the rise in a given distance to the right, the larger the positive

value of the slope), and the slope of any line that *falls* as it goes from left to right will be negative (the greater the fall in a given distance to the right, the greater the negative value of the slope).

Table 20.12

PRODUCTION AND WASTE FOR EIGHT PUNCH-PRESS OPERATORS, WITH CORRESPONDING Z-SCORES OF THE PRODUCTION AND WASTE FIGURES

Operator	*Production*	*Waste*	*z-Score in production*	*z-Score in waste*
1	95	3.0	−.69	−1.38
2	103	4.5	+.31	+.82
3	88	3.5	−1.56	−.65
4	98	4.0	−.31	+.09
5	93	3.0	−.94	−1.38
6	107	4.5	+.81	+.82
7	114	4.0	+1.69	+.09
8	106	5.0	+.69	+1.56
Mean	100.5	3.94		
S.D.	8.0	.68		

By line of *best fit* in the definition is meant a line so drawn that the sum of the squared deviations in a vertical direction from the original points to the line is less than the sum would be for any other straight line that might be drawn.

A rough approximation of the value of r may be obtained by plotting the z-scores of the two variables, fitting a straight line to these points by inspection, and graphically measuring the slope of this straight line. Although this method is never used in practical computation (because it is both inaccurate and laborious), the application of it to a set of representative data may serve to clarify the meaning of the correlation coefficient, r. Returning to the data for which we have previously computed the rank-order correlation (see Table 20.11 on p. 638), we first compute the z-scores for each measure as shown in Table 20.12.

These pairs of z-scores are used as the x and y values for 8 points that are ploted on co-ordinate axes as in Fig. 20.6. The straight line that seems best to fit these points is then determined (as with a stretched string that is moved about until the desired location is obtained) and drawn on the graph. The correlation, r, as determined by this crude method, is obtained by measuring the slope of this line. The procedure applied to Fig. 20.6 gives a value of $r = .61$, but it should be emphasized that this value is affected by:

1. The accuracy with which the straight line has been located, and
2. The accuracy with which the slope of the line has been measured after it has been drawn.

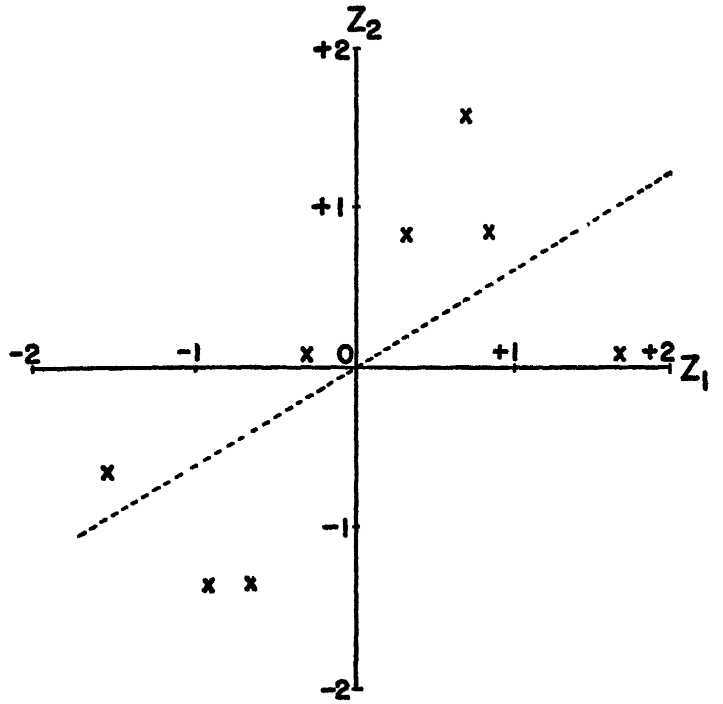


Fig. 20.6. A plot of the z-scores for production and waste records of the eight punch-press operators shown above.

Points (1) and (2) both operate to eliminate the possibility of complete accuracy in this method of determining a correlation coefficient. Therefore, a mathematical method has been devised to make the computation, so that no plotting of points or graphic measurements are required. This method involves determining the equation of the straight line that, if plotted, would best fit the points, and computing the slope of this straight line of best fit from the equation.

It may be proved mathematically that the slope of the straight line of best fit is given by the following equation[4]: Slope $= r = \dfrac{\Sigma Z_x Z_y}{N}$,

[4] The proof of this formula is as follows:

$Z_{x1}, Z_{x2}, Z_{x3}, \ldots \quad Z_{xn}$ are the z-scores for the x variable
$Z_{y1}, Z_{y2}, Z_{y3}, \ldots \quad Z_{yn}$ are the z-scores for the y variable

The equation of any straight line is: $y = a + bx$. The problem is to find the values of the constants a and b in this equation that will give the straight line that "best fits" the data according to the criterion of best fit stated on p. 640, that is, the straight line

where $\Sigma Z_x Z_y$ is read "the sum of the products of the z-scores for the pairs of points or values."

Applying the formula to the data in Table 20.11, we may compute the correlation as in Table 20.13.

Table 20.13

COMPUTATION OF r BY Z-SCORE METHOD BETWEEN PRODUCTION AND WASTE FIGURES

Operator	*Production*	*Waste*	*z-Score in production*	*z-Score in waste*	$(z_1 z_2)$
1	95	3.0	−.69	−1.38	+.95
2	103	4.5	+.31	+.82	+.25
3	88	3.5	−1.56	−.65	+1.01
4	98	4.0	−.31	+.09	−.03
5	93	3.0	−.94	−1.38	+1.30
6	107	4.5	+.81	+.82	+.66
7	114	4.0	+1.69	+.09	+.15
8	106	5.0	+.69	+1.56	+1.08
					+5.38

$$r = \frac{Z_x Z_y}{\mathrm{N}} = \frac{+5.38}{8} = .67$$

that will give a minimum value to the sum of the squared deviations between the line and the original points.

The first point (whose co-ordinate points are Z_{x1} and Z_{y1} will deviate from the line by an amount:

$$d_1 = Z_{y1} - a - bZ_{x1}$$

this deviation squared will be

$$d_1{}^2 = (Z_{y1} - a - bZ_{x1})^2$$

If the sum of all such squared deviations of the points from the line be represented by u, then

$$u = \Sigma d^2 = \Sigma\,(Z_y - a - bZ_x)^2$$

For the values of a and b that result in u being a minimum, the derivatives of u with respect to a and b, respectively, must be zero. To find the correlation coefficient it is therefore necessary only to differentiate the equation with respect to a and to b, to set the resulting derivatives equal to zero, and to solve for b (which is the slope of the straight line of best fit). This is done as follows:

$$\frac{\partial u}{\partial a} = 0 = 2\,\Sigma(Z_y - a - bZ_x)\,(-1) \tag{1}$$

$$\frac{\partial u}{\partial b} = 0 = 2\,\Sigma(Z_y - a - bZ_x)\,(-Z_x) \tag{2}$$

The value of r thus obtained by computation, .67, differs from the value of .61 obtained by plotting and inspection. The plotting and inspection method yielded a value that was somewhat in error for the data in question.

---

(1) above becomes:

$$0 = -\Sigma Z_y + Na + b\Sigma Z_x$$
$$\Sigma Z_y = Na + b\Sigma Z_x \tag{3}$$

(2) above becomes:

$$0 = -\Sigma Z_x Z_y + a\Sigma Z_x + b\Sigma Z_x^2$$
$$\Sigma Z_x Z_y = a\Sigma Z_x + b\Sigma Z_x^2 \tag{4}$$

It will be remembered that a z-score is obtained as follows (see p. 630):

$$Z_x = \frac{X - M_x}{\sigma_x}$$

Where X is a given raw score, M_x is the mean raw score, and σ_x the standard deviation of the raw scores.

The sum of all z-scores is therefore:

$$\begin{aligned}\Sigma Z_x &= \frac{\Sigma(X - M_x)}{\sigma_x}\\ &= \frac{\Sigma X}{\sigma_x} - \frac{NM_x}{\sigma_x}\\ &= \frac{\Sigma X}{\sigma_x} - \frac{\frac{N\Sigma X}{N}}{\sigma_x}\\ &= \frac{\Sigma X}{\sigma_x} - \frac{\Sigma X}{\sigma_x}\\ &= 0\end{aligned}$$

In a similar way it can be shown that

$$\Sigma Z_y = 0$$

By substituting 0 for ΣZ_x and ΣZ_y in (3) we find immediately that a in the equation of the straight line of best fit is zero.

Working with (4), we find the value of b (which is the slope or the correlation coefficient) as follows:

$$\Sigma Z_x Z_y = a\Sigma Z_x + b\Sigma Z_x^2$$

since $\Sigma Z_x = 0$, this becomes:

$$\Sigma Z_x Z_y = b\Sigma Z_x^2$$
$$b = \frac{\Sigma Z_x Z_y}{\Sigma Z_x^2} \tag{5}$$

It may be shown as follows that $\Sigma Z_x^2 = N$

$$Z_x = \frac{X - M_x}{\sigma_x}$$

While the z-score method of computing a correlation coefficient illustrated in Table 20.13 may be used with any number of cases and will yield the correct mathematical value of r, the use of this method when many pairs of data are to be correlated is very laborious. It is therefore recommended, under such circumstances, that a modification of the fundamental formula $r = \frac{\Sigma Z_x Z_y}{N}$, which makes it possible to compute r from raw score values rather than z-score values, be used. One

---

$$Z_x^2 = \frac{(X - M_x)^2}{\sigma_x^2}$$

$$\Sigma Z_x^2 = \frac{\Sigma (X - M_x)^2}{\sigma_x^2}$$

$$\Sigma Z_x^2 = \frac{\Sigma (X^2 - 2XM_x + M_x^2)}{\sigma_x^2}$$

$$\Sigma Z_x^2 = \frac{\Sigma X^2 - 2M_x \Sigma X + NM_x^2}{\sigma_x^2}$$

$$\Sigma Z_x^2 = \frac{\Sigma X^2 - 2\frac{\Sigma X}{N}\Sigma X + N\left(\frac{\Sigma X}{N}\right)^2}{\sigma_x^2}$$

$$\Sigma Z_x^2 = \frac{\dfrac{N\Sigma X^2 - 2(\Sigma X)^2 + (\Sigma X)^2}{N}}{\dfrac{\Sigma X^2}{N} - \left(\dfrac{\Sigma X}{N}\right)^2}$$

$$\Sigma Z_x^2 = \frac{\dfrac{N\Sigma X^2 - (\Sigma X)^2}{N}}{\dfrac{N\Sigma X^2 - (\Sigma X)^2}{N^2}}$$

$$\Sigma Z_x^2 = \frac{N^2(\Sigma X^2) - N(\Sigma X)^2}{N\Sigma X^2 - (\Sigma X)^2}$$

$$\Sigma Z_x^2 = \frac{N[N\Sigma X^2 - (\Sigma X)^2]}{[N\Sigma X^2 - (\Sigma X)^2]}$$

$$\Sigma Z_x^2 = N \qquad (6)$$

Substituting the value of $\Sigma\ (Z_x)^2$ given in (6) in equation (5), we have the slope or $r = b = \frac{\Sigma Z_x Z_y}{N}$.

convenient formula[5] for determining the coefficient of correlation directly from the raw data is:

$$r = \frac{N\Sigma XY - \Sigma X\Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

[5] The proof of this formula is as follows:

$$r = \frac{\Sigma Z_x Z_y}{N}$$

$$= \frac{1}{N} \sum \frac{(X - M_x)}{\sigma_x} \frac{(Y - M_y)}{\sigma_y}$$

$$= \frac{1}{N} \sum \frac{(XY - XM_y - YM_x + M_x M_y)}{\sigma_x \sigma_y}$$

$$= \frac{1}{N} \sum \frac{\left(XY - \frac{X\Sigma Y}{N} - Y\frac{\Sigma X}{N} + \frac{\Sigma X\Sigma Y}{N^2}\right)}{\sigma_x \sigma_y}$$

$$= \frac{1}{N} \frac{\Sigma XY - \frac{\Sigma X\Sigma Y}{N} - \frac{\Sigma X\Sigma Y}{N} + \frac{N\Sigma X\Sigma Y}{N^2}}{\sqrt{\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2}\sqrt{\frac{\Sigma Y^2}{N} - \left(\frac{\Sigma Y}{N}\right)^2}}$$

$$= \frac{1}{N} \frac{\frac{N\Sigma XY - \Sigma X\Sigma Y}{N}}{\sqrt{\frac{N\Sigma X^2 - (\Sigma X)^2}{N^2}}\sqrt{\frac{N\Sigma Y^2 - (\Sigma Y)^2}{N^2}}}$$

$$= \frac{1}{N} \frac{\frac{N\Sigma XY - \Sigma X\Sigma Y}{N}}{\frac{\sqrt{N\Sigma X^2 - (\Sigma X)^2}}{N}\ \frac{\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}{N}}$$

$$= \frac{1}{N} \frac{N^2(N\Sigma XY - \Sigma X\Sigma Y)}{N\sqrt{N\Sigma X^2 - (\Sigma X)^2}\ \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

$$= \frac{N\Sigma XY - \Sigma X\Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\ \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

When we apply this formula, for illustrative purposes, to the data tabulated in Table 20.13, the computations shown in Table 20.14 result:

Table 20.14

COMPUTATION OF *r* DIRECTLY FROM RAW DATA

Operator	*Production* (*X*)	*Waste* (*Y*)	X^2	Y^2	XY
1	95	3.0	9,025	9.00	285.0
2	103	4.5	10,609	20.25	463.5
3	88	3.5	7,744	12.25	308.0
4	98	4.0	9,604	16.00	392.0
5	93	3.0	8,649	9.00	279.0
6	107	4.5	11,449	20.25	481.5
7	114	4.0	12,996	16.00	456.0
8	106	5.0	11,236	25.00	530.0
	$\Sigma X = 804$	$\Sigma Y = 31.5$	$\Sigma X^2 = 81{,}312$	$\Sigma Y^2 = 127.75$	$\Sigma XY = 3{,}195.0$

$$r = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\ \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

$$r = \frac{8(3{,}195) - (804)\,(31.5)}{\sqrt{8(81{,}312) - (804)^2}\ \sqrt{8(127.75) - (31.5)^2}}$$

$$r = .67$$

When a considerable number of pairs of data are to be correlated, the use of a chart will still further simplify the computations. Several forms of such a chart have been prepared. One convenient form is shown in Fig. 20.7. This chart shows the computation of the correlation between time used in inspecting 300 pieces of material and the number of defective pieces detected. In using this chart the following steps should be followed:

1. Decide upon appropriate class intervals for one of the variables (using the rules given on p. 617) and write these in on either the *x* or the *y* axis.
2. Decide upon appropriate class intervals for the other variables and write these in on the axis not used in (1) above.
3. Place 1 tally mark on the scattergram for each pair of values being correlated. For example, if an inspector spotted 33 defects in 16.5 minutes, the tally mark would go in the cell that is found at the intersection of the *row* containing 33 defects and the *column* containing 16.5 minutes.
4. After all tally marks have been placed on the chart, the rows should be added horizontally and the sum of the tally marks in each row written opposite this row in the f_y column (Column *1*).

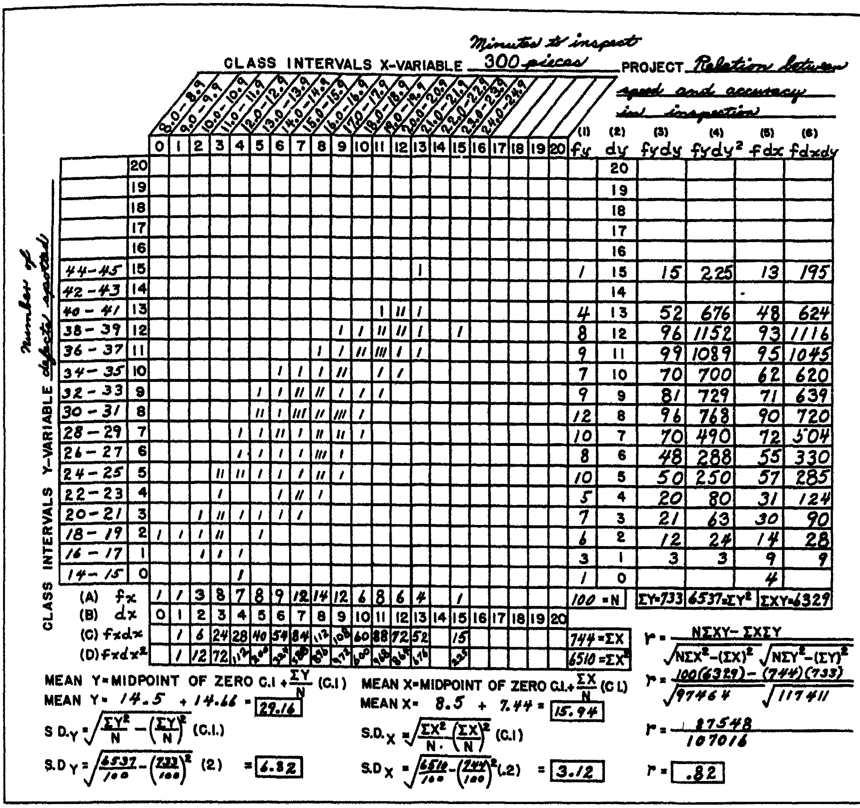


Fig. 20.7. A chart used in the computation of a product moment coefficient of correlation.

5. The tally marks in each column should be added and the sum written at the bottom of each column in the row (*A*), the f_x row.
6. The f_y column should be added and the sum written opposite N at the bottom of this column. *The value of* N *thus obtained may be checked by adding the values in the* f_x *row. The sum of these values should also give the value of* N.
7. Each value of f_y in the column so headed should be multiplied by the value of d_y opposite it, and the resultant product written in Column *3*, headed f_yd_y. *The sum of Column 3 is the value of* ΣY *which is used in the formula.*
8. Each value in Column *3*, the $f_y\,d_y$ column, should be multiplied by the corresponding value in Column *2*, the d_y column, resulting in the values for Column *4*, or the $f_yd_y^2$ column. *The sum of Column 4 is the value of* ΣY^2 *which is used in the formula.*
9. The values going into Column *5*, the fd_x column, are determined by finding, for each row, the sum of the products of the number of cases in each cell times the *x* value of that cell. For example, in the first

row in which a tally mark appears, there is only a single case, which appears in the cell under an x value of 13. The value to go into the blank in Column *5* is therefore (1) (13) = 13. In the next row no tally marks appear; therefore, this row is blank. In the next row, 1 tally mark appears in the cell under an x value of 11, 2 in the cell with an x value of 12, and 1 in the cell with an x value of 13. The value to go into the blank cell in Column *5* is therefore (1) (11) + (2) (12) + (1) (13) = 48. The remaining cells in Column *5* are filled in a similar manner.

10. The cells in Column *6*, the fd_xd_y column, are filled with values obtained by multiplying each value in Column *2*, the d_y column, by the value in that same row appearing in Column *5*, the fd_x column. The value in the first cell in Column *6* is therefore (15) (13) = 195. *The sum of Column 6 is the value of* ΣXY *which is used in the formula.*
11. The values in row (*C*) are obtained by multiplying each value in row (*A*), the f_x row, by the value directly below it in row (*B*), the d_x row. The values appearing in row *A* have already been obtained (see Step 5 above). The resultant values are entered in row (*C*), the f_xd_x row. *The sum of the values appearing in row (C) is the value of* ΣX *which is used in the formula.*
12. Each value in row (*B*), the d_x row, should be multiplied by the value directly below in row (*C*), the f_xd_x row. The resultant values should be entered in row (*D*), the $f_xd_x^2$ row. *The sum of the values in row (D) is the value of* ΣX^2 *which is used in the formula.*
13. The values for N (see Step *6*), ΣY (see Step *7*), ΣY^2 (see Step *8*), ΣXY (see Step *10*), ΣX (see Step *11*), and ΣX^2 (see Step *12*) are now entered in the formula. The indicated arithmetic computations are then performed, yielding the value of r.

The use of this method assumes that each measure has the value of the midpoint of the class interval in which it falls. The computations indicated on the chart result in obtaining not only the value for r but also the Mean and the Standard Deviations of both the X and Y arrays. It will be noted that these are the same formulas previously considered on pp. 625 and 628.

Appendix B

Taylor-Russell Tables[1]

(For Institutional Prediction)

TABLES OF THE PROPORTION WHO WILL BE SATISFACTORY AMONG THOSE SELECTED FOR GIVEN VALUES OF THE PROPORTION OF PRESENT EMPLOYEES CONSIDERED SATISFACTORY, THE SELECTION RATIO, AND R

Proportion of Employees Considered Satisfactory = .05
Selection Ratio

r	.05	.10	.20	.30	40	.50	.60	.70	.80	.90	.95
.00	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05
.05	.06	.06	.06	.06	.06	.05	.05	.05	.05	.05	.05
.10	.07	.07	.07	.06	.06	.06	.06	.05	.05	.05	.05
.15	.09	.08	.07	.07	.07	.06	.06	.06	.05	.05	.05
.20	.11	.09	.08	.08	.07	.07	.06	.06	.06	.05	.05
.25	.12	.11	.09	.08	.08	.07	.07	.06	.06	.05	.05
.30	.14	.12	.10	.09	.08	.07	.07	.06	.06	.05	.05
.35	.17	.14	.11	.10	.09	.08	.07	.06	.06	.05	.05
.40	.19	.16	.12	.10	.09	.08	.07	.07	.06	.05	.05
.45	.22	.17	.13	.11	.10	.08	.08	.07	.06	.06	.05
.50	.24	.19	.15	.12	.10	.09	.08	.07	.06	.06	.05
.55	.28	.22	.16	.13	.11	.09	.08	.07	.06	.06	.05
.60	.31	.24	.17	.13	.11	.09	.08	.07	.06	.06	.05
.65	.35	.26	.18	.14	.11	.10	.08	.07	.06	.06	.05
.70	.39	.29	.20	.15	.12	.10	.08	.07	.06	.06	.05
.75	.44	.32	.21	.15	.12	.10	.08	.07	.06	.06	.05
.80	.50	.35	.22	.16	.12	.10	.08	.07	.06	.06	.05
.85	.56	.39	.23	.16	.12	.10	.08	.07	.06	.06	.05
.90	.64	.43	.24	.17	.13	.10	.08	.07	.06	.06	.05
.95	.73	.47	.25	.17	.13	.10	.08	.07	.06	.06	.05
1.00	1.00	.50	.25	.17	.13	.10	.08	.07	.06	.06	.05

[1] *These tables are reproduced by permission from H. C. Taylor and J T. Russell, "The Relationship of Validity Coefficients to the Practical Effectiveness of Tests in Selection: Discussion and Tables,"* Journal of Applied Psychology, *23 (1939), 565–578.*

Proportion of Employees Considered Satisfactory = .10
Selection Ratio

r	.05	.10	.20	.30	.40	50	.60	.70	.80	90	.95
.00	.10	.10	.10	.10	.10	.10	.10	.10	.10	10	.10
.05	.12	.12	.11	.11	.11	.11	.11	.10	.10	.10	.10
.10	.14	.13	.13	.12	.12	.11	.11	.11	.11	.10	10
.15	.16	.15	.14	.13	.13	12	.12	.11	.11	.10	.10
.20	.19	.17	.15	.14	.14	13	.12	.12	.11	.11	.10
.25	.22	.19	.17	.16	.14	.13	13	.12	.11	.11	.10
.30	.25	.22	.19	.17	.15	.14	.13	.12	.12	.11	10
.35	.28	.24	.20	.18	.16	.15	.14	.13	.12	.11	.10
40	.31	.27	.22	.19	.17	.16	.14	.13	.12	.11	.10
.45	.35	.29	.24	.20	.18	.16	.15	.13	.12	.11	.10
.50	.39	.32	.26	.22	.19	.17	.15	.13	.12	.11	.11
.55	.43	.36	.28	.23	.20	.17	.15	.14	.12	.11	.11
60	.48	.39	.30	.25	.21	.18	.16	.14	.12	.11	.11
.65	.53	.43	.32	.26	.22	.18	.16	.14	.12	.11	.11
.70	.58	47	.35	.27	.22	.19	.16	.14	.12	.11	.11
75	.64	.51	.37	.29	.23	.19	.16	.14	.12	.11	.11
.80	.71	.56	.40	.30	.24	.20	.17	.14	.12	.11	.11
.85	.78	.62	.43	.31	.25	.20	.17	.14	.12	.11	.11
.90	.86	.69	.46	.33	.25	.20	.17	.14	.12	.11	.11
.95	.95	.78	.49	.33	.25	.20	.17	14	.12	.11	.11
1.00	1.00	1.00	50	.33	.25	.20	.17	.14	.13	11	.11

Proportion of Employees Considered Satisfactory = .20
Selection Ratio

r	.05	.10	20	.30	.40	.50	.60	.70	.80	.90	.95
.00	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20
.05	.23	.23	.22	.22	.21	.21	.21	.21	.20	.20	.20
.10	.26	.25	.24	.23	.23	.22	22	.21	.21	.21	.20
.15	.30	.28	.26	.25	.24	.23	.23	.22	.21	.21	.20
.20	.33	.31	.28	.27	.26	.25	.24	.23	.22	.21	.21
.25	.37	.34	.31	.29	.27	.26	.24	.23	.22	.21	.21
.30	.41	.37	.33	.30	.28	.27	.25	.24	.23	.21	.21
.35	.45	.41	.36	.32	.30	.28	.26	.24	.23	.22	.21
.40	.49	.44	.38	.34	.31	.29	.27	.25	.23	.22	.21
.45	.54	.48	.41	.36	.33	.30	.28	.26	.24	.22	.21
.50	.59	.52	.44	.38	.35	.31	.29	.26	.24	.22	.21
.55	.63	.56	.47	.41	.36	.32	.29	.27	.24	.22	.21
.60	.68	.60	.50	.43	.38	.34	.30	.27	.24	.22	.21
.65	.73	.64	.53	.45	.39	.35	.31	.27	.25	.22	.21
.70	.79	.69	.56	.48	.41	.36	.31	.28	.25	.22	.21
.75	.84	.74	.60	.50	.43	.37	.32	.28	.25	.22	.21
.80	.89	.79	.64	.53	.45	.38	.33	.28	.25	.22	.21
.85	.94	.85	.69	.56	.47	.39	33	.28	.25	.22	.21
.90	.98	.91	.75	.60	.48	.40	.33	.29	.25	.22	.21
.95	1.00	.97	.82	.64	.50	.40	33	.29	.25	.22	.21
1.00	1.00	1.00	1.00	.67	.50	.40	.33	.29	.25	.22	.21

Proportion of Employees Considered Satisfactory = .30
Selection Ratio

r	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
.00	.30	.30	.30	.30	.30	.30	.30	.30	.30	.30	.30
.05	.34	.33	.33	.32	.32	.31	.31	.31	.31	.30	.30
.10	.38	.36	.35	.34	.33	.33	.32	.32	.31	.31	.30
.15	.42	.40	.38	.36	.35	.34	.33	.33	.32	.31	.31
.20	.46	.43	.40	.38	.37	.36	.34	.33	.32	.31	.31
.25	.50	.47	.43	.41	.39	.37	.36	.34	.33	.32	.31
.30	.54	.50	.46	.43	.40	.38	.37	.35	.33	.32	.31
.35	.58	.54	.49	.45	.42	.40	.38	.36	.34	.32	.31
.40	.63	.58	.51	.47	.44	.41	.39	.37	.34	32	.31
.45	.67	.61	.55	.50	.46	.43	.40	.37	.35	.32	.31
.50	.72	.65	.58	.52	.48	.44	.41	.38	.35	.33	.31
.55	.76	.69	.61	.55	.50	.46	.42	.39	.36	.33	.31
.60	.81	.74	.64	.58	.52	.47	.43	.40	.36	.33	.31
.65	.85	.78	.68	.60	.54	.49	.44	.40	.37	.33	.32
.70	.89	.82	.72	.63	.57	.51	.46	.41	.37	.33	.32
.75	.93	.86	.76	.67	.59	.52	.47	.42	.37	.33	.32
.80	.96	.90	.80	.70	.62	.54	.48	.42	.37	.33	.32
.85	.99	.94	.85	.74	.65	.56	.49	.43	.37	.33	.32
.90	1.00	.98	90	.79	.68	.58	.49	.43	.37	.33	.32
.95	1.00	1.00	.96	.85	.72	.60	.50	.43	.37	.33	.32
1.00	1.00	1.00	1.00	1.00	.75	.60	.50	.43	.38	.33	.32

Proportion of Employees Considered Satisfactory = .40
Selection Ratio

r	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
.00	.40	.40	.40	.40	.40	.40	.40	.40	.40	.40	.40
.05	.44	.43	.43	.42	.42	.42	.41	.41	.41	.40	.40
.10	.48	.47	.46	.45	.44	.43	.42	.42	.41	.41	.40
.15	.52	.50	.48	.47	.46	.45	.44	.43	.42	.41	.41
.20	.57	.54	.51	.49	.48	.46	.45	.44	.43	.41	.41
.25	.61	.58	.54	.51	.49	.48	.46	.45	.43	.42	.41
.30	.65	.61	.57	.54	.51	.49	.47	.46	.44	.42	.41
.35	.69	.65	.60	.56	.53	.51	.49	.47	.45	.42	.41
.40	.73	.69	.63	.59	.56	.53	.50	.48	.45	.43	.41
.45	.77	.72	.66	.61	.58	.54	.51	.49	.46	.43	.42
.50	.81	.76	.69	.64	.60	.56	.53	.49	.46	.43	.42
.55	.85	.79	.72	.67	.62	.58	.54	.50	.47	.44	.42
.60	.89	.83	.75	.69	.64	.60	.55	.51	.48	.44	.42
.65	.92	.87	.79	.72	.67	.62	.57	.52	.48	.44	.42
.70	.95	.90	.82	.76	.69	.64	.58	.53	.49	.44	.42
.75	.97	.93	.86	.79	.72	.66	.60	.54	.49	.44	.42
.80	.99	.96	.89	.82	.75	.68	.61	.55	.49	.44	.42
.85	1.00	.98	.93	.86	.79	.71	.63	.56	.50	.44	.42
.90	1.00	1.00	.97	.91	.82	.74	.65	.57	.50	.44	.42
.95	1.00	1.00	.99	.96	.87	.77	.66	.57	.50	.44	42
1.00	1.00	1.00	1.00	1.00	1.00	.80	.67	.57	.50	.44	.42

Proportion of Employees Considered Satisfactory = .50
Selection Ratio

r	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
.00	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50
.05	.54	.54	.53	.52	.52	.52	.51	.51	.51	.50	.50
.10	.58	.57	.56	.55	.54	.53	.53	.52	.51	.51	.50
.15	.63	.61	.58	.57	.56	.55	.54	.53	.52	.51	.51
.20	.67	.64	.61	.59	.58	.56	.55	.54	.53	.52	.51
.25	.70	.67	.64	.62	.60	.58	.56	.55	.54	.52	.51
.30	.74	.71	.67	.64	.62	.60	.58	.56	.54	.52	.51
.35	.78	.74	.70	.66	.64	.61	.59	.57	.55	.53	.51
.40	.82	.78	.73	.69	.66	.63	.61	.58	.56	.53	.52
.45	.85	.81	.75	.71	.68	.65	.62	.59	.56	.53	.52
.50	.88	.84	.78	.74	.70	.67	.63	.60	.57	.54	.52
.55	.91	.87	.81	.76	.72	.69	.65	.61	.58	.54	.52
.60	.94	.90	.84	.79	.75	.70	.66	.62	.59	.54	.52
.65	.96	92	.87	.82	.77	.73	.68	.64	.59	.55	.52
.70	.98	.95	.90	.85	.80	.75	.70	.65	.60	.55	.53
.75	.99	.97	.92	.87	.82	.77	.72	.66	.61	.55	.53
.80	1.00	.99	.95	.90	.85	.80	.73	.67	.61	.55	.53
.85	1.00	.99	.97	.94	.88	.82	.76	.69	.62	.55	.53
.90	1.00	1.00	.99	.97	.92	.86	.78	.70	.62	.56	.53
.95	1.00	1.00	1.00	.99	.96	.90	.81	.71	.63	.56	.53
1.00	1.00	1.00	1.00	1.00	1.00	1.00	.83	.71	.63	.56	.53

Proportion of Employees Considered Satisfactory = .60
Selection Ratio

r	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
.00	.60	.60	.60	.60	.60	.60	.60	.60	.60	.60	.60
.05	.64	.63	.63	.62	.62	.62	.61	.61	.61	.60	.60
.10	.68	.67	.65	.64	.64	.63	.63	.62	.61	.61	.60
.15	.71	.70	.68	.67	.66	.65	.64	.63	.62	.61	.61
.20	.75	.73	.71	.69	.67	.66	.65	.64	.63	.62	.61
.25	.78	.76	.73	.71	.69	.68	.66	.65	.63	.62	.61
.30	.82	.79	.76	.73	.71	.69	.68	.66	.64	.62	.61
.35	.85	.82	.78	.75	.73	.71	.69	.67	.65	.63	.62
.40	.88	.85	.81	.78	.75	.73	.70	.68	.66	.63	.62
.45	.90	.87	.83	.80	.77	.74	.72	.69	.66	.64	.62
.50	.93	.90	.86	.82	.79	.76	.73	.70	.67	.64	.62
.55	.95	.92	.88	.84	.81	.78	.75	.71	.68	.64	.62
.60	.96	.94	.90	.87	.83	.80	.76	.73	.69	.65	.63
.65	.98	.96	.92	.89	.85	.82	.78	.74	.70	.65	.63
.70	.99	.97	.94	.91	.87	.84	.80	.75	.71	.66	.63
.75	.99	.99	.96	.93	.90	.86	.81	.77	.71	.66	.63
.80	1.00	.99	.98	.95	.92	.88	.83	.78	.72	.66	.63
.85	1.00	1.00	.99	.97	.95	.91	.86	.80	.73	.66	.63
.90	1.00	1.00	1.00	.99	.97	.94	.88	.82	.74	.67	.63
.95	1.00	1.00	1 00	1.00	.99	.97	.92	.84	.75	.67	.63
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.86	.75	.67	.63

Proportion of Employees Considered Satisfactory = .70
Selection Ratio

r	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
.00	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70
.05	.73	.73	.72	.72	.72	.71	.71	.71	.71	.70	.70
.10	.77	.76	.75	.74	.73	.73	.72	.72	.71	.71	.70
.15	.80	.79	.77	.76	.75	.74	.73	.73	.72	.71	.71
.20	.83	.81	.79	.78	.77	.76	.75	.74	.73	.71	.71
.25	.86	.84	.81	.80	.78	.77	.76	.75	.73	.72	.71
.30	.88	.86	.84	.82	.80	.78	.77	.75	.74	.72	.71
.35	.91	.89	.86	.83	.82	.80	.78	.76	.75	.73	.71
.40	.93	.91	.88	.85	.83	.81	.79	.77	.75	.73	.72
.45	.94	.93	.90	.87	.85	.83	.81	.78	.76	.73	.72
.50	.96	.94	.91	.89	.87	.84	.82	.80	.77	.74	.72
.55	.97	.96	.93	.91	.88	.86	.83	.81	.78	.74	.72
.60	.98	.97	.95	.92	.90	.87	.85	.82	.79	.75	.73
.65	.99	.98	.96	.94	.92	.89	.86	.83	.80	.75	.73
.70	1.00	.99	.97	.96	.93	.91	.88	.84	.80	.76	.73
.75	1.00	1.00	.98	.97	.95	.92	.89	.86	.81	.76	.73
.80	1.00	1.00	.99	.98	.97	.94	.91	.87	.82	.77	.73
.85	1.00	1.00	1.00	.99	.98	.96	.93	.89	.84	.77	.74
.90	1.00	1.00	1.00	1.00	.99	.98	.95	.91	.85	.78	.74
.95	1.00	1.00	1.00	1.00	1.00	.99	.98	.94	.86	.78	.74
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.88	.78	.74

Proportion of Employees Considered Satisfactory = .80
Selection Ratio

r	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
.00	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80
.05	.83	.82	.82	.82	.81	.81	.81	.81	.81	.80	.80
.10	.85	.85	.84	.83	.83	.82	.82	.81	.81	.81	.80
.15	.88	.87	.86	.85	.84	.83	.83	.82	.82	.81	.81
.20	.90	.89	.87	.86	.85	.84	.84	.83	.82	.81	.81
.25	.92	.91	.89	.88	.87	.86	.85	.84	.83	.82	.81
.30	.94	.92	.90	.89	.88	.87	.86	.84	.83	.82	.81
.35	.95	.94	.92	.90	.89	.89	.87	.85	.84	.82	.81
.40	.96	.95	.93	.92	.90	.89	.88	.86	.85	.83	.82
.45	.97	.96	.95	.93	.92	.90	.89	.87	.85	.83	.82
.50	.98	.97	.96	.94	.93	.91	.90	.88	.86	.84	.82
.55	.99	.98	.97	.95	.94	.92	.91	.89	.87	.84	.82
.60	.99	.99	.98	.96	.95	.94	.92	.90	.87	.84	.83
.65	1.00	.99	.98	.97	.96	.95	.93	.91	.88	.85	.83
.70	1.00	1.00	.99	.98	.97	.96	.94	.92	.89	.85	.83
.75	1.00	1.00	1.00	.99	.98	.97	.95	.93	.90	.86	.83
.80	1.00	1.00	1.00	1.00	.99	.98	.96	.94	.91	.87	.84
.85	1.00	1.00	1.00	1.00	1.00	.99	.98	.96	.92	.87	.84
.90	1.00	1.00	1.00	1.00	1.00	1.00	.99	.97	.94	.88	.84
.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.99	.96	.89	.84
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.89	.84

Proportion of Employees Considered Satisfactory = .90
Selection Ratio

r	.05	.10	.20	30	.40	.50	.60	.70	.80	.90	.95
.00	.90	90	.90	.90	.90	90	.90	.90	.90	90	.90
.05	.92	91	91	91	.91	.91	.91	90	.90	.90	.90
.10	.93	.93	.92	92	.92	.91	.91	.91	.91	.90	.90
.15	95	.94	.93	.93	.92	.92	.92	.91	.91	91	.90
20	.96	.95	.94	.94	.93	.93	.92	.92	.91	91	.90
.25	.97	.96	.95	.95	.94	.93	.93	.92	.92	.91	.91
.30	.98	.97	.96	.95	.95	.94	.94	.93	.92	.91	.91
.35	.98	.98	.97	.96	.95	95	.94	93	.93	.92	91
.40	.99	.98	.98	.97	96	.95	.95	.94	.93	.92	.91
.45	.99	.99	.98	.98	.97	.96	.95	.94	.93	92	91
.50	1.00	.99	.99	.98	.97	.97	.96	.95	.94	.92	.92
.55	1.00	1.00	.99	.99	.98	.97	.97	96	.94	93	.92
.60	1.00	1.00	99	.99	.99	.98	.97	.96	.95	93	.92
.65	1.00	1 00	1.00	.99	99	.98	.98	.97	.96	.94	.92
.70	1.00	1.00	1.00	1.00	.99	.99	.98	.97	.96	.94	.93
.75	1.00	1.00	1.00	1 00	1.00	.99	.99	.98	.97	.95	.93
.80	1 00	1.00	1.00	1 00	1 00	1.00	.99	.99	.97	.95	.93
.85	1.00	1 00	1 00	1.00	1 00	1 00	1.00	.99	.98	.96	.94
.90	1.00	1 00	1.00	1.00	1.00	1 00	1.00	1.00	.99	.97	.94
.95	1.00	1.00	1.00	1 00	1 00	1.00	1.00	1.00	1 00	.98	.94
1 00	1.00	1.00	1.00	1 00	1.00	1.00	1.00	1.00	1.00	1.00	.95

Appendix C

Lawshe Expectancy Tables[1] (For Individual Prediction)

Per Cent of Employees Considered Satisfactory = 30%

r	INDIVIDUAL PREDICTOR CATEGORIES				
	Hi 1/5	Next 1/5	Middle 1/5	Next 1/5	Lo 1/5
.15	38	32	30	28	22
.20	40	34	29	26	21
.25	43	35	29	24	19
.30	46	35	29	24	16
.35	49	36	29	22	14
.40	51	37	28	21	12
.45	55	38	28	20	10
.50	58	38	27	18	09
.55	61	39	27	17	07
.60	64	40	26	15	05
.65	68	41	25	13	04
.70	72	42	23	11	03
.75	76	43	22	09	02
.80	80	44	20	06	01
.85	85	45	17	04	00
.90	90	46	12	02	00
.95	96	48	07	00	00

[1] *C. H. Lawshe, R. A. Bolda, R. L. Brune, and G. Auclair, "Expectancy Charts III, Their Theoretical Development,"* Personnel Psychology, *11 (1958), 545–599.*

Per Cent of Employees Considered Satisfactory = 40%

r	INDIVIDUAL PREDICTOR CATEGORIES				
	Hi 1/5	Next 1/5	Middle 1/5	Next 1/5	Lo 1/5
.15	48	44	40	36	32
.20	51	45	40	35	30
.25	54	44	40	34	28
.30	57	46	40	33	24
.35	60	47	39	32	22
.40	63	48	39	31	19
.45	66	49	39	29	17
.50	69	50	39	28	14
.55	72	53	38	26	12
.60	75	53	38	24	10
.65	79	55	37	22	08
.70	82	58	36	19	06
.75	86	59	35	17	04
.80	89	61	34	14	02
.85	93	64	32	10	01
.90	97	69	29	06	00
.95	100	76	23	02	00

Per Cent of Employees Considered Satisfactory = 50%

r	INDIVIDUAL PREDICTOR CATEGORIES				
	Hi 1/5	Next 1/5	Middle 1/5	Next 1/5	Lo 1/5
.15	58	54	50	46	42
.20	61	55	50	45	39
.25	64	56	50	44	36
.30	67	57	50	43	33
.35	70	58	50	42	30
.40	73	59	50	41	28
.45	75	60	50	40	25
.50	78	62	50	38	22
.55	81	64	50	36	19
.60	84	65	50	35	16
.65	87	67	50	33	13
.70	90	70	50	30	10
.75	92	72	50	28	08
.80	95	75	50	25	05
.85	97	80	50	20	03
.90	99	85	50	15	01
.95	100	93	50	08	00

Per Cent of Employees Considered Satisfactory = 60%

r	INDIVIDUAL PREDICTOR CATEGORIES				
	Hi 1/5	Next 1/5	Middle 1/5	Next 1/5	Lo 1/5
.15	68	64	60	57	52
.20	71	63	60	56	48
.25	73	65	60	55	48
.30	76	66	61	54	44
.35	78	68	61	53	40
.40	81	69	61	52	37
.45	83	71	61	51	34
.50	86	72	62	50	31
.55	88	74	62	48	28
.60	90	76	62	47	25
.65	92	78	63	45	21
.70	94	80	64	43	18
.75	96	83	65	42	14
.80	98	86	66	39	11
.85	99	90	68	36	07
.90	100	94	71	31	03
.95	100	98	77	24	00

Per Cent of Employees Considered Satisfactory = 70%

r	INDIVIDUAL PREDICTOR CATEGORIES				
	Hi 1/5	Next 1/5	Middle 1/5	Next 1/5	Lo 1/5
.15	77	73	69	69	62
.20	79	75	70	67	59
.25	81	75	71	65	58
.30	84	76	71	65	54
.35	86	78	71	64	52
.40	88	79	72	63	49
.45	90	80	72	63	46
.50	91	82	73	62	42
.55	93	83	73	61	39
.60	95	85	74	60	36
.65	96	87	75	59	32
.70	97	89	77	58	29
.75	98	91	78	57	25
.80	99	94	80	56	20
.85	100	96	83	55	16
.90	100	98	88	54	10
.95	100	100	93	52	04

Appendix D

Publishers of Representative Personnel Tests

The tests in this Appendix include only a very small number of the personnel tests commercially available. They have been limited almost entirely to tests that have been discussed in this book. Very good additional sources of information about the availability and uses of personnel tests are:

O. K. Buros, ed., *The Fifth Mental Measurements Yearbook* (New Brunswick, N. J.: Rutgers University Press, 1959).

This book, in addition to giving publishers and prices, contains critical reviews of each test by one or more competent authorities.

R. M. Dorcus and M. H. Jones, *Handbook of Employee Selection* (New York: McGraw-Hill Book Company, 1950).

This book gives abstracts of 426 articles dealing with personnel tests that have been published in professional journals. It contains indexes by job, test, and author. The abstracts are factual and non-critical.

In the lists given below, the publisher of each test is coded in the column indicated.

Title	*Code to publisher* *(see pp. 662–663)*
Intelligence Tests	
Adaptability Test	13
Army Alpha, Forms A and B, Bregman's Revision	10
Army Alpha, Forms 5 and 7, Well's Revision	10
Army Alpha, Short Form 6	10
Army General Classification Test, First Civilian Edition	13
Benge Employment Tests	9
California Test of Mental Maturity	3
Cattell Culture-free Test	10
Chicago Non-Verbal Examination	10
D.A.T. Abstract Reasoning Test	10

D.A.T. Numerical Reasoning Test	10
D.A.T. Verbal Reasoning Test	10
Ohio State University Psychological Test, Form 21	13
Oral Directions Test: An Intelligence Test on Phonograph Records	10
O'Rourke General Classification Test	11
Otis Self-Administering Tests of Mental Ability	10
Psychological Examination of the American Council on Education	4
Purdue Non-Language Personnel Test	13
Revised Beta Examination (non-verbal)	10
S.R.A. Non-Verbal Form	13
S.R.A. Primary Mental Abilities Tests	13
S.R.A. Verbal Form	13
Thurstone Test of Mental Alertness	13
Wonderlic Personnel Test	10

Clerical Tests

E.R.C. Stenographic Aptitude Test	13
D.A.T. Clerical Speed and Accuracy	10
General Clerical Test	10
Minnesota Clerical Test (formerly Minnesota Vocational Test for Clerical Workers)	10
Purdue Clerical Adaptability Test Form A (revised)	19
S.R.A. Clerical Aptitudes	13

Industrial Vision Tests

Keystone Visual Safety Tests	7
Ortho-Rater	2
Sight-Screener	1
Vision Tester	17

Stenographic and Typing Skills

Bennett Stenographic Aptitude Test	10
Bennett Stenographic Proficiency Test	10
Blackstone Stenographic Proficiency Tests	18
Seashore-Bennett Stenographic Proficiency Test (on phonograph records)	10
S.R.A. Typing Skills	13
S.R.A. Dictation Skills	13
Thurstone Examination in Typing	10

Dexterity and Manipulative Tests

MacQuarrie Test for Mechanical Ability	3
Minnesota Rate of Manipulation Test	16

O'Connor Finger Dexterity Test	8
O'Connor Tweezer Dexterity Test	8
Purdue Hand Precision Test	8
Purdue Grooved Pegboard	8
Purdue Pegboard	13

Mechanical Tests

Bennett Mechanical Comprehension Test	10
Industrial Training Classification Test	19
Minnesota Mechanical Assembly Test	15
Minnesota Paper Form Board, revised	10
Minnesota Spatial Relations Test	10
O'Connor-Wiggly Block Test	5
O'Rourke Mechanical Aptitude Test—Jr. Grade	10
Purdue Mechanical Adaptability Test	19
Purdue Mechanical Performance Test	8
S.R.A. Mechanical Aptitudes	12
Stenquist Mechanical Aptitude Test	10

Personality Tests

A-S Reaction Study, Beckman's Revision for Business Use	10
Gordon Personal Profile	10
Gordon Personal Inventory	10
Bell Adjustment Inventory	10
Bernreuter Personality Inventory	15
Edwards Personal Preference Schedule	10
Guilford Inventory of Factors GAMIN	14
Guilford Inventory of Factors STDCR	14
Guilford-Martin Personnel Inventory	14
Guilford-Zimmerman Temperament Survey	14
Humm-Wadsworth Temperament Scale	6
Minnesota Multiphasic Personality Inventory	10
Rorschach Projective Test	10
Thurstone Temperament Schedule	13

Interest Tests

Cardall Primary Business Interests Tests	13
Cleeton Vocational Interest Inventory	10
Kuder Preference Record	13
Strong Vocational Interest Blank for Men, Form M	10
Strong Vocational Interest Blank for Women, Form W	10
Study of Values	10

Trade Tests

Purdue Industrial Mathematics Test	19
Purdue Interview Aids	
Can You Read a Working Drawing?	19
Can You Read a Micrometer?	19
Can You Read a Scale?	19
Purdue Vocational Tests	
Technical Information in Machine Shop	19
Technical Information in Electricity	19
Blueprint Reading Test	19
Purdue Trade Information Test in Welding	19
Purdue Trade Information Test in Carpentry	19
Purdue Trade Information Test in Engine Lathe Operation	19
Purdue Trade Information Test for Sheetmetal Workers	19

General

Purdue Reading Test for Industrial Supervisors	19
Purdue Industrial Supervisors Word-Meaning Test	19
Purdue Creativity Test	19

Industrial Batteries

Flanagan Industrial Tests	13
Employee Aptitude Survey	12

PUBLISHERS

1. American Optical Company	Southbridge, Massachusetts
2. Bausch & Lomb Optical Company	Rochester, New York
3. California Test Bureau	5916 Hollywood Boulevard Los Angeles 28, California
4. Educational Testing Service	2½ Chambers Street Princeton, New Jersey
5. Human Engineering Laboratory, Inc.	347 Beacon Street Boston 16, Massachusetts
6. Humm Personnel Service	P.O. Box 1433 Del Valle Station Los Angeles 15, California
7. Keystone View Company	Meadville, Pennsylvania
8. Lafayette Instrument Company	52 By-Pass Lafayette, Indiana
9. Management Service Company	3136 North 24th Street Philadelphia, Pennsylvania

10.	Psychological Corporation	304 East 45th Street New York 17, New York
11.	Psychological Institute	P.O. Box 1117 Lake Alfred, Florida
12.	Psychological Services, Inc.	1800 Wilshire Blvd. Los Angeles 57, California
13.	Science Research Associates	259 East Erie Street Chicago 11, Illinois
14.	Sheridan Supply Company	P.O. Box 837 Beverly Hills, California
15.	Stanford University Press	Stanford, California
16.	C. H. Stoelting Company	424 North Homan Avenue Chicago 20, Illinois
17.	Titmus Optical Company	Petersburg, Virginia
18.	Harcourt Brace & World, Inc.	757 Third Ave. New York 17, New York
19.	University Book Store	360 State Street West Lafayette, Indiana

Index

Index

GEORGE ALLEN & UNWIN LTD

Head Office
40 Museum Street, London, W.C.1
Telephone: 01-405 8577

Sales, Distribution and Accounts Departments
Park Lane, Hemel Hempstead, Herts.
Telephone: 0442 3244

Athens: 34 Panepistimiou Street
Auckland: P.O. Box 36013, Northcote Central, N.4
Barbados: P.O. Box 222, Bridgetown
Beirut: Deeb Building, Jeanne d'Arc Street
Bombay: 103/5 Fort Street, Bombay 1
Buenos Aires: Escritorio 454-459, Florida 165
Calcutta: 285 J Bepin Behari, Ganguli Street, Calcutta 12
Cape Town: 68 Shortmarket Street
Hong Kong: 105 Wing On Mansion, 26 Hancow Road, Kowloon
Ibadan: P.O. Box 62
Karachi: Karachi Chambers, McLeod Road
Madras: 2/18 Mount Road, Madras
Mexico: Villalongin 32, Mexico 5, D.F.
Nairobi: P.O. Box 30583
Philippines: P.O. Box 157, Quezon City D-502
Rio de Janeiro: Caixa Postal 2537-Zc-00
Singapore: 36c Prinsep Street, Singapore 7
Sydney, N.S.W.: Bradbury House, 55 York Street
Tokyo: C.P.O. Box 1728, Tokyo 100-91
Toronto: 81 Curlew Drive, Don Mills